CHAPTER IX

Theories of Magnetic Storms and Auroras

1. Birkeland, K., *Arch. Sci. Phys.*, **1**, 497 (1896).
2. Goldstein, S., *Wiedemanns Ann.*, **12**, 266 (1881).
3. Birkeland, K., *Vidensk. Skrifter, I. Mat. Naturv. kl.* 1901, Christiania, 1901.
4. Birkeland, K., *Norwegian Aurora Polaris Expedition*, 1902-03, Christiania, **1**, Part 1 (1908); Part 2 (1913).
5. Störmer, C., *Arch. Sci. Phys. Genève*, (*a*) **24**, 5, 113, 221, 317 (1907); (*b*) **32**, 33, (1911, 32); (*c*) **35**, 483 (1913).
6. Störmer, C., *Phys. Rev.*, **45**, 835 (1934).
7. Schuster, A., *Proc. Roy. Soc. A*, **85**, 44 (1911).
8. Lindemann, F. A., *Phil. Mag.*, **38**, 669 (1919).
9. Chapman, S., *Proc. Phil. Soc.*, **21**, 577 (1923).
10. Chapman, S. and Ferraro, V. C. A., *Terr. Mag. and Atmos. Elec.*, **36**, 77, 171 (1931); **37**, 147 (1932).
11. Alfvén, H., 'A theory of Magnetic storms and of the Aurorae,' *Kungl. Sve. Vetenskaps Akad. Handl., Stockholm, III*, **18**, No. 3 (1939); *ibid.*, **18**, No. 9 (1940).
12. Martyn, D. F., *Nature*, **167**, 92 (1951).
13. Pettit, E., *Publications of Yerkes Observatory*, **3**, 205 (1925); *Astrophys. Jour.*, **76**, 9 (1932); *ibid.*, **84**, 319 (1936).
14. Milne, E. A., *Mon. Not. Roy. Astr. Soc.*, **86**, 459 and 578 (1926); *Handbuch der Astrophysik*, **3**, 1. *Halfte*, pp. 173–183, J. Springer, Berlin (1930).
15. Alfvén, H., *Zeit. Phys.*, **105**, 633 (1937).
16. Hale, G. E., *Astrophys. Jour.*, **38**, 27 (1913) and 99 (1919); see also **47**, 206 (1918).
17. Thiessen, G., *Zeit. Astrophys.*, **26**, 16 (1949).
18. Saha, M. N., *Observatory*, **66**, 13 (1945).
19. Richardson, R. S., *Trans. Amer. Geophys. Union*, p. 558 (1944).

19*a*. Brück, H. A. and Ruttlant, F., *Mon. Not. Roy. Astr. Soc.*, **106**, 130 (1946).

20. Chapman, S., *Mon. Not. Roy. Astr. Soc.*, **89**, 456 (1929).

20*a*. Gartlein, G. W., *Nature*, **167**, 277 (1951).

20*b*. Meinel, A. B., *Astrophys. Jour.*, **113**, 50 (1951).

21. Maunder, E. W., *Mon. Not. Roy. Astr. Soc.*, **76**, 63 (1913).

21*a*. Chapman, S. and Bartels, J., *Geomagnetism*, Vol. II, p. 804, Oxford University Press (1940).

22. Chree, C. and Stagg, J. M., *Phil. Trans. Roy. Soc. A*, **227**, 21 (1927); **212**, 75 (1912).
23. Cortie, A. L., *Mon. Not. Roy. Astr. Soc.*, **73**, 52 (1913).
24. Bartels, J., *Terr. Mag. and Atmos. Elec.*, **37**, 1 (1932).
25. Brüche, E., *Terr. Mag. and Atmos. Elec.*, **36**, 41 (1931); *Naturwiss.*, **18**, 1085 (1930); *Phys. Zeit.*, **31**, 1011 (1930); **32**, 31 (1931); *Zeit. Astrophys.*, **2**, 30 (1931).
26. Brüche, E., *Terr. Mag. and Atmos. Elec.*, **36**, 41 (1931).
27. Malmfors, K. G., *Ark. Mat. Astr. Fys.*, **34B**, No. 1 (1946).
28. Chapman, S., 'Emission Spectra of the Night Sky and Aurorae', *Gassiot Committee Report*, p. 131 (1947), published by the Physical Society, London (1948).
29. Cowling, T. G., *Terr. Mag. and Atmos. Elec.*, **47**, 209 (1942).
30. Chapman, S., 'Emission Spectra of the Night Sky and Aurorae', *Gassiot Committee Report*, p. 128 (1947), published by the Physical Society, London (1948).
31. Paulsen, A., *Sur la nature et l' origine de l' aurore boreale* (Kjöbenhavn, 1894).
32. Paulsen, A., *Sur les récentes théories de l' aurore polaire* (*Kjöbenhavn*), *Vid. Selk. Overs.*, No. 2, 109 (1906).
33. Hulburt, E. O., *Terr. Mag. and Atmos. Elec.*, **33**, 11 (1928); *Phy. Rev.*, **31**, 1038 (1928).
34. Chapman, S., *Phys. Rev.*, **32**, 993 (1928).

THE UPPER ATMOSPHERE

S. K. MITRA

SIR RASHBEHARY GHOSE PROFESSOR OF PHYSICS
UNIVERSITY OF CALCUTTA

SECOND EDITION

THE ASIATIC SOCIETY
1 PARK STREET, CALCUTTA 16

THE UPPER ATMOSPHERE

S. K. MITRA.

MONOGRAPH SERIES No. V

1st Edition, January, 1947
2nd Edition, January, 1952
Reprint of 2nd Edition, March, 1992

Published by
Dr. Chandan Roy Chaudhuri
General Secretary
The Asiatic Society
1, Park Street
Calcutta 700016

Printed by
Jayanta Bagchi
P. M. Bagchi & Company (P) Limited
19, Gulu Ostagar Lane
Calcutta 700006.

Price :

Hard Bound	Rs.	250.00
	$	50.00
Paperback	Rs.	150.00
	$	40.00

TO

SIR EDWARD APPLETON

G.B.E., K.C.B., M.A., D.Sc., LL.D., F.R.S.,

NOBEL LAUREATE

PREFATORY NOTE

The publication of the book, "The Upper Atmosphere", by Prof. S. K. Mitra, a major work in contemporary science was an important event in the history of publication of the Asiatic Society. Prof Meghnad Saha, the then President of the Society strongly recommended its publication when some publishers were reluctant to take it up. Consequently the book came out in 1947 and was immediately hailed as a pioneering work by the world scientific community. A second edition came out in 1952. In spite of its demand the book was not reprinted for a long time.

After the rejuvenation of the Society in 1984, preparations were on to reprint the volume before 1989, the birth centenary year of Prof. S. K. Mitra. However, this was not materialized. When the present Council took charge in March, 1992 it gave top priority to the project and is extremely happy to bring out this volume which, we believe, will be an asset both to the students as well as to the researchers. The subject of upper atmosphere has undergone a sea change in the last forty years. This has been explained in the foreword by Prof. A. P. Mitra, FRS, who is also a former pupil of Prof. S. K. Mitra.

20 March, 1992.
Calcutta

Dr. Chandan Roy Chaudhuri
General Secretary

FOREWORD

The beginning of radioscience in India was made some hundred years ago during the turn of the last century by Sir J. C. Bose through a remarkable series of experiments on generation, detection and characterisation of microwaves over 25mm to 5mm, long before study of microwaves became fashionable.

After that there was a lull for over twenty years, until the establishment of Wireless Laboratory. Serious work on ionosphere, however, began only around 1930, when Professor S. K. Mitra with H. Rakshit, using a medium wave transmitter made available by the Calcutta station of the Indian State Broadcasting service, the first experimental evidence of the E-region of the ionosphere was obtained in India. This was achieved in 1930 (published in Nature) only a few years after the discovery of the ionosphere by Appleton and Barnett (by using a loop and a vertical aerial receiving the direct and sky wave in 1925) and the pulse experiment of Breit and Tuve (1926, the forerunner of the ionosonde, later to be extensively used for routine monitoring of the ionosphere). A series of papers came in quick succession relating to the behaviour of ionospheric layers over Calcutta. It was remarkable that the quality of ionospheric mapping achieved at that time with such simple equipments was so excellent.

This was an exciting period for ionospheric science in India. The timing was also right. The Second Polar Year Programme (IPY2) was being organized, and a study of the Ionosphere was being included for the first time. Mitra decided to formally participate in this programme: this was India's first entry into organized international science. Along with experimental observations of the ionospheric layers, questions were being asked about the origin of these layers. That the primary agency is solar radiation was accepted — the mechanism for a layer structure with a specific maximum was also understood following the work of Chapman (1931), as the combined effect of decreasing density and of the increasing intervals of radiation with height. Chapman's theory was based on a single ionizable constituent and a monochromatic radiation. The question was the nature of the constituent (or constituents) and the wavelengths of the radiations. There was also no clear ideas about the relative contributions (if any) of energetic particles from the sun and the electromagnetic radiations. There was certainly evidences of ionization produced by solar particles in the higher magnetic latitudes as one would expect for charged particles under influence of the terrestrial magnetic field, but at low latitudes the effects were yet to be adequately assessed. There was other ionizing sources: cosmic rays, meteors, thunderstorms that needed to be examined. An excellent opportunity soon came that allowed a distinction between ionization produced by solar electromagnetic radiation and solar corpuscles. This was the occurrence of the annular solar eclipse visible in Calcutta on August 21, 1933.

Another unusual experiment (for that time) was a conscious effort to look for ionization caused by meteor showers. Earlier efforts by Schafer and Goodall looking for changes in the E-region equivalent height due to Leonid meteor shower of 1931 were unfortunately marred by the occurrence of a magnetic storm. Mitra and his colleagues decided to use the opportunity of Leonid meteor shower of 1933 and, instead of looking for changes in equivalent height, decided to measure directly changes in ionization through changes in critical frequency.

Scientists were already looking for existence of additional layers in the Ionosphere. There was indirect evidence of a layer below the E-layer at levels where collisions

between electrons and ions and gaseous molecules are frequent and so absorption of medium wave radio waves is large. Increase of atmospherics and of medium wave signal strength have been noted. Mitra and Shyam announced in 1935 the detection of regular echoes from low heights (55 km) ; later even echoes as low as 20 km (20-30 km range) were reported. The former were attributed to D-region, and the very low level reflections were believed to come from a hitherto unsuspected layer, that Mitra called "C-layer".

Reflections from heights around and above 55 km were later to be observed extensively with HF transmissions, particularly in Australia, USA and Canada, and the technique came to be known in later years as Partial Reflection Technique. The very low level reflections that Mitra and his colleagues detected from heights of 20-30 km and also reported soon after by Coldwell and Friend in the USA and Watson-Watt, Bainbridge-Bell, Wilkins and Bowell in U. K. were, however, not taken seriously for a long time until the concept of using HF atmospheric radars came up in the sixties. In that sense these early pioneering works can be treated as the forerunner of HF radars.

We have to understand all this background to realise why Mitra decided in 1935 to write a detailed status report on the Ionosphere and the emerging new mode of radio communication through the Ionosphere. Following a symposium organized by the National Institute of Sciences which he opened, Mitra prepared a comprehensive report on the state of knowledge of the Ionosphere available at that time. This was published in the Proceedings of the National Institute of Sciences. The wider concept of the upper atmosphere (the neutral atmosphere, motions, airglow, chemistry) was missing in the report.

"The Upper Atmosphere", first published by the Asiatic Society, in 1947 and in revised form in 1952 was a major international event. Professor J. N. Bhar, in the biographical memoir of S. K. Mitra, pointed out some of the intriguing features of the publication of this book. It was his intention to have the book published by a foreign publisher, and he wrote to two or three firms, but the replies were disappointing and in some ways curious. To quote from the reply from one of the publishers:

"From previous experience of books of this nature we feel very doubtful whether it would have a large enough sale to cover the expense of publication: in fact we anticipate that it would involve us in a considerable financial loss. A further consideration is that even in the small field covered by your book it would have to compete with Chapman and Bartels' Geomagnetism and works by Sir Napier Shaw". And yet when the book was published by the Asiatic Society, 2000 copies were sold out within three years, the second revised edition published in 1952 went out of print in a few years, and generations of students in radio communication, in ionospheric and upper atmospheric physics, in geomagnetism, in space science have been using this book as a major reference document.

We came to Professor Mitra's laboratory when the second edition was under preparation. As postgraduate students our first duties were to take turns in running the ionospheric recorder which was then under the overall control of Dr. S. S. Baral. There were several Ph.D. students: M. K. Dasgupta, R.B. Banerjee, A. K. Saha, Miss Mrinmayi Ghosh and myself. The work of revision of the different chapters of the first edition and of assistance in the new chapters or sections, such as those on escape of gases from terrestrial atmosphere, atmospheric tides, rocket exploration of the upper atmosphere, were considered as excellent training for research.

"The Upper Atmosphere" was in many ways different from 1935 report, although partially based on it. The most important difference was that it considered for the first time the atmospheric environment as a whole, neutral and ionized, its thermal structure and distribution of constituents, its motions, the interaction of the solar

radiation and particle stream with these gaseous constituents, and the mechanism of airglow. The Ionosphere was treated as only a part of this vast panorama that interlinked in a myriad ways the sun, the earth and the atmosphere. This was then an entirely new concept. Secondly, deviating from the then existing practice of studying the Ionosphere from the point of view of propagation of radio waves, he viewed the exploring radiowave as a remote sensing tool sensing levels which could not be reached with balloons and were only beginning to be explored by rockets. The entire chapter on temperature in the upper atmosphere (Chapter XI), for example, discussed at length the various indirect methods of estimating temperature in the upper atmosphere: from considerations of escape of helium, observations of auroral streamers at heights of 1000 km, twilight flash of the red oxygen lines traceable upon 1300 km, from a number of ionospheric parameters such as electron collision frequency, scale heights, recombination coefficient; from the width of emission lines from the night airglow. From the helium escape he concluded that the 600-800 km, the temperature should be of the order of 1000°K; from the observed twilight and auroral luminisity effects a temperature of about 1100°K, from collision frequencies temperatures around 3000°K at 400 km, from auroral scale heights around 2700°K and from electron concentrations around 1500-2500°K. Although such values were necessarily gross and the large differences between the different estimates indicated that, the main point was the deduction of high temperatures in the upper atmosphere. In the section of Concluding Remarks he comes up with models of atmospheric density and temperature which have remarkable similarities with contemporary models using satellite drag. In his address as President of INSA in 1960, he brings this point up as in the following way:

"It may be noted that the result of high density in the highest atmospheric regions referred to above is not quite unexpected. It confirms the assumption of a rising temperature in these regions as has always been made in constructing atmospheric models. In Table IV are given the densities and temperatures at some selected heights of such a model constructed some years age on certain assumptions (Mitra, 1952). These may be compared with the corresponding figures in the model shown in Table III prepared recently by A. P. Mitra and S.B. Mathur (1959) from high altitude rocket and satellite data.

Table III : A Model upper atmosphere based on height

Height (km)	Density (gm/cm^3)	Pressure (mm of Hg)	Temperature (°K)
800	6.00×10^{-17}	4.8×10^{-10}	2068
500	1.54×10^{-15}	9.46×10^{-9}	1591
300	4.97×10^{-14}	1.85×10^{-7}	1023
130	274×10^{-11}	3.46×10^{-5}	462

Table IV : An older Model (S.K. Mitra's Upper Atmosphere)

Height (km)	Density (gm/cm^3)	Pressure (mm of Hg)	Temperature (°K)
800	3.67×10^{-17}	2.13×10^{-10}	3040
500	7.00×10^{-16}	2.25×10^{-9}	1840
300	2.11×10^{-14}	3.83×10^{-8}	1040
110	4.98×10^{-10}	2.70×10^{-4}	270
100	1.74×10^{-9}	1.00×10^{-3}	240

In both the models there is significant density even at 800 km height".

I should mention at this stage that at the time of launching of Sputnik I in 1957, the Russian space scientists found that the only reasonable atmospheric models they could use for predicting the lifetimes of satellites were those given in the "Upper Atmosphere".

Ionospheric chemistry, as we know now, had not yet emerged but even here the book made a beginning through detailed discussions of the formation and destruction of ozone, of dissociation of N_2 and O_2 and of night airglow. For ozone the descriptions were surprisingly detailed; there were also discussions of the heat balance of the stratosphere including cooling and heating of the middle atmosphere due to ozone absorption -- problems that we are still discussing in essentially similar manner.

The last edition was published exactly four decades age. During this period the entire picture of the Sun-Earth system had drastically changed. The upper boundary of the atmospheric environment has now been pushed to many earth radii i.e. to hundreds of thousands of kilometres with different components of these atmospheric environment interacting with each other. In this vastly changed (and greatly expanded) canvas, much of the information given in this book will be considered out of date, and values quoted sometimes substantially different from those accepted now. Furthermore, new areas of special interest, such as the effects of human activities on global environment (eg. ozone holes, high power radio wave heating, effects of spacecraft effluents) do not naturally find any place here.

Nevertheless four generations of atmospheric scientists have grown with this book and have drawn inspiration and often a new insight. Those whose scientific activities spanned the decades of the 50s, 60s and 70s have without exception consulted this book in some form or other. And remarkably the book has been read by people with widely different interests -- by telecommunication engineers, by meteorologists, by upper atmospheric scientists, by global environmental experts, by solar terrestrial specialists and even those concerned with policy matters.

It continues to be used as a reference book in advanced courses on atmospheric physics in all parts of the world. Many have indeed expressed disappointment that copies are no longer available.

We believe that the book with its vast canvas and clear exposition of the basic concepts of interrelationship between the sun and the different components of the atmospheric environment will still be of considerable value. This is particularly important now in the context of the International Geosphere Biosphere Programme and global change activities.

The question of revising the book substantially to make it more contemporary was discussed at various times. There was also a proposal that some of his erstwhile students could take up this job of revision. All these would have taken time, and would also have diminished the historical value of the book. Considering all these aspects, the book is being reprinted without any change.

Bhatnagar Fellow
National Physical Laboratory
Dr. K. S. Krishnan Road
New Delhi 110012
March 24, 1992

A. P. MITRA
Former Director General
C.S.I.R., Government of India

PREFACE TO THE SECOND EDITION

The first edition of the book was sold out within an unexpectedly short time. As the demand for the book persisted the preparation of a second edition was undertaken at the request of the publishers, the Asiatic Society, and also of many readers in different parts of the world. The work was begun in March, 1950, and the original plan was to have it completed by the beginning of 1951. Unfortunately, various unforeseen circumstances (amongst which my enforced absence from the country for some time was one) delayed the publication by nearly a year.

Since the publication of the first edition results of new upper atmospheric researches have been appearing at a very rapid rate. As such, additions and revisions much more extensive than originally contemplated, had to be made in the preparation of the new edition. The more important of these are mentioned below.

In Chapter I the section dealing with the escape of the atmospheric gases has been revised in the light of the recent work on the subject by Lyman Spitzer, Jr.

A non-mathematical treatment of the mechanism of the atmospheric tides is given in Chapter II after Weekes and Wilkes.

In Chapter IIIA, Öpik's theory of meteors is given more prominence (after a simplified treatment by Herlofson). The theory of Lindemann and Dobson, though it has been subjected to criticism, is retained on account of its lasting value. Results of radar study of meteors, as far as have been available up to the time of writing, have been included.

In Chapter IV the results of newer studies on the variations of the ozone content with geographic latitude and with tropospheric conditions have been added.

Chapter V in the first edition dealt only with the dissociation of molecular oxygen in the upper atmosphere. To this has now been added the dissociation of molecular nitrogen. The analysis of the dissociation and distribution-in-height of molecular oxygen is as given in the first edition. Criticisms of this analysis, as also of the results obtained thereof, have, however, been made recently by H. E. Moses and Ta-You Wu [*Phys. Rev.*, **83**, 109 (1951)]. Unfortunately, reference to this work could not be made as the Chapter in question had already been printed off. In the discussion on the dissociation of molecular nitrogen, reference to a prior work of G. Hertzberg and L. Hertzberg [*Nature*, **161**, 283 (1948)] on the subject has been omitted through inadvertence. But, this is referred to in Chapter VIII.

In Chapter VI the various methods of sounding the ionosphere by means of radio waves have been reviewed in greater detail. The theories

of the formation of the various ionospheric regions have been revised according to current ideas. The limitations of the geometrical optics method for studying electromagnetic wave propagation in the ionospheric regions are referred to in a separate section. A new section on the estimation of recombination coefficient and rate of electron production from ionospheric data has been added. The section on radio fade-out has been extended by inclusion of the phenomenon of 'Sudden-Phase-Anomaly' (S.P.A.). A fuller account is given of the geomagnetic control of the F-layer. A new section on tides (Martyn's theory) and travelling disturbances in the ionosphere has been added. The subjects 'Scattering in the Ionosphere' and 'Luxembourg Effect' are dealt with more fully.

In Chapter VII the large S_q- and L- variations at Huancayo, near the geomagnetic equator, are specially delineated after Bartels. The Huancayo anomaly is also referred to in Chapter XIII. The probable current systems for an individual magnetic storm, as computed from magnetic observatories data, are represented after Vestine.

The list of lines and bands in the auroral spectrum (Chap. VIII) and the theories of their modes of excitation are revised and made up-to-date. Evidence of the entry of protons in the upper atmosphere is referred to in this connection.

In Chapter IX the major additions have been accounts of Alfvén's theory (though it has been subjected to criticisms) and of the extension of the Chapman-Ferraro theory by Martyn. The Birkeland-Störmer theory, notwithstanding the fundamental objections against it, is retained. This is because, the analysis of the motion of a charged particle in the field of a magnetic dipole, as made by these authors, has many applications, irrespective of the correctness of the theory.

In Chapter X the contemporary Russian work on zodiacal light and on gegenschein phenomena are discussed. In the list of the lines and bands of the night air-glow spectrum the newly identified OH (Meinel) bands have been included. The discussion on the probable modes of excitations of the lines and bands is revised. Data of the results of height measurements of the luminescent layers are collected in a Table. A new section on hydrogen in the upper atmosphere has been added. The section on sodium in the upper atmosphere is amplified by inclusion of results of more recent studies.

Chapter XI on the temperature distribution in the ionospheric regions is revised after the recent discussion on the subject by Gerson.

A new chapter on the rocket explorations of the upper atmosphere (Chapter XII) has been added. Results of studies by V-2 and Aerobee flights, as far as available up to the time of writing, have been included.

In Chapter XIII, the surveys of the contemporary state of our knowledge and of the unsolved problems of the upper atmosphere are revised in the light of the later observations and data. The subject of winds in the high regions of the upper atmosphere is briefly discussed.

In the Appendix, besides some minor revisions, a new section on the proposed upper atmospheric nomenclature has been added.

Notwithstanding the additions and revisions as listed above, I am acutely conscious of the fact that there have been many omissions. This is due partly to my ignorance and partly to the accumulation of facts and data too rapid to be properly systematized. I shall be grateful if the readers of the book bring to my notice instances of errors and omissions so that they may be rectified in a future edition.

I shall feel amply compensated for the trouble of preparing this second edition, if it is received by the readers with the same cordiality as the first edition.

January, 1952, S. K. M.

Institute of Radio Physics and Electronics,
University College of Science, Calcutta.

PREFACE TO THE FIRST EDITION

The present volume on the *Upper Atmosphere* had a modest beginning. In August, 1935, I had the honour of being invited by the National Institute of Sciences of India to open a symposium on the ionosphere. I took the opportunity to write out my address in the form of a report on the *Present State of Our Knowledge of the Ionosphere.* This was published in the *Proceedings* of the Institute (Vol. I, pp. 131ff.) and was favourably received by ionospheric workers in different parts of the world. This emboldened me to think of attempting a treatise on the upper atmosphere —of which the ionosphere is only one of the phases—as I found that in spite of the many and varied contributions to its study by theoretical and solar physicists, by geomagneticians, by meteorologists and by others, there hardly existed a book dealing with the subject as a whole and in a comprehensive manner. I confided my idea to my friend and colleague. Professor M. N. Saha (then Professor of Physics at the University of Allahabad) while both of us were on a visit to the United Kingdom in 1936. Professor Saha enthusiastically supported the idea and I planned the preparation of the book on my return to India in the fall of the same year. Unfortunately, the intervention of the war and various other causes interrupted the progress of the work and also delayed its publication. This has now been made possible through the active interest which the Royal Asiatic Society of Bengal took in the work after Professor Saha became its President (1945-46). I take the opportunity of thanking the Society for undertaking the publication at a time, when, though the war was over, conditions were far from normal.

During the period that elapsed between the planning and the publication of the book (precisely 11 years—one complete solar cycle!) two excellent treatises, one on *Terrestrial Magnetism and Electricity*, edited by Dr. J. A. Fleming and the other on *Geomagnetism* (Vols. I and II), by Prof. S. Chapman and Dr. J. Bartels, appeared. Being specialized treatises, these books discuss all the aspects of geomagnetism in great detail including those which influence or are influenced by the upper atmospheric phenomena. The present volume can hardly claim the same thoroughness and wealth of detail in dealing with these particular topics. It is hoped, however, that the surveys of these subjects as given here will be helpful to the reader in understanding the general nature of the phenomena and their relationship with other atmospheric problems.

It has been my aim to give in the book, in a connected form, an account of the present state of our knowledge of the upper atmosphere as has been obtained so far by direct and indirect observations. It is, however, impossible in these days of intense specialized research and rapid advancement of knowledge, to make a volume like this fully up-to-date. New

facts and new results are emerging fast. And, one has to resist strongly the temptation of including the latest developments if one intends to give a finality to his project. Nevertheless, I greatly deplore my inability to include, at least, the results of experiments in the U.S.A. with the V-2 rockets, news about which reached India just as the printing of the book was nearing completion.

No attempt has been made to make the references in the bibliography complete. Indeed, for a volume like this, this is well-nigh an impossible task. The references are representative rather than exhaustive; but wherever they relate directly to the subject-matter under discussion, they have been freely given.

A large amount of knowledge—a surprisingly large amount considering the inaccessibility of the regions concerned—has already been gained of the upper atmospheric regions. But, as the perusal of the volume will show, much still remains obscure and speculative. My labour in writing the book will be amply rewarded if its study stimulates further investigations on the yet unsolved problems.

January, 1947, S. K. M.

University College of Science,

Calcutta.

ACKNOWLEDGEMENTS

As for the first edition, I received for the preparation of this second edition, ungrudging help from my colleagues and from research workers in my laboratory.

Dr. J. N. Bhar, Reader in the Department of Radio Physics and Electronics was always ready with his help and advice on matters concerning the printing work.

Dr. S. N. Ghosh, Imperial Chemical Industries Research Fellow, helped in the revision of Chapter II.

Mr. S. S. Baral, Senior Research Assistant, helped in revising Chapter VI. He also shouldered a large part of the responsibility of carrying on correspondence with the press.

Mr. A. P. Mitra, Junior Research Assistant, prepared portions of Chapters VI and XI and the Appendix. He also rendered valuable help by checking data and drafting texts of new matters whenever required to do so.

Mr. A. K. Saha, Junior Research Assistant, helped in the revisions of Chapters VII, VIII and IX. A number of new diagrams in these and in some other chapters were also drawn by him.

Mr. R. B. Banerji, Government of India Research Scholar, and Mr. M. K. Das Gupta, Research Assistant, helped in revising the subject matters of some of the chapters. Miss Mrinmoyee Ghosh, Ghose Research Scholar, besides rendering similar help, undertook the arduous task of compiling the Index of Subjects. Mr. R. K. Mitra, Junior Research Assistant, helped in the preparation of the Index of Authors. Mr. D. K. Bose, Ghose Research Scholar, prepared a number of diagrams.

I would also take this opportunity of putting on record my grateful thanks to Professor S. Chapman for advising me regarding certain revisions; to Dr. D. F. Martyn for bringing to my notice certain inaccurate statements in the first edition and for going through the subject-matters of Sec. 14(*c*) of Chapter VI and Sec. 4(*c*) of Chapter IX; to Dr. F. L. Whipple for inviting attention to a doubtful opinion expressed regarding sporadic meteors in Chapter IIIB in the first edition and supplying original photograph for Fig. 3 in the same Chapter; to Professor A. C. B. Lovell for making available original photographs for Figs. 13, 15, 16 and 21 in Chapter IIIB; to Dr. D. R. Bates for information concerning contemporary ideas on the excitation processes of night air-glow; to Dr. F. E. Roach for discussion by correspondence on certain aspects of twilight-flash spectra; to Dr. J. Bartels for inviting attention to the Huancayo anomaly in Chapter VII, Sec. 4.

As in the first edition, a large number of diagrams have been taken or adapted from articles published in standard scientific journals. Due acknowledgement to the author is made in every case.

Finally, I must thank the Asiatic Society for undertaking the publication of the present edition in spite of the very large increase in the printing cost and also for the full freedom they gave me not only in respect of revising the text-matter, but also in that of adding to the numbers of Plates and diagrams.

CONTENTS

CHAPTER I

GENERAL CONSIDERATIONS

CHAPTER II

ATMOSPHERIC OSCILLATIONS—TIDES IN THE HIGH ATMOSPHERE

CHAPTER III

TEMPERATURE AND DENSITY DISTRIBUTION IN THE MIDDLE ATMOSPHERE

IIIA—Anomalous Propagation of Sound

CHAPTER IV

THE OZONOSPHERE

CHAPTER V

Oxygen and Nitrogen in the Upper Atmosphere

CHAPTER VI

The Ionosphere

CHAPTER VII

Electrical Currents in the Upper Atmosphere: Terrestrial Magnetic Variations

CHAPTER VIII

Aurora Polaris

CHAPTER IX

Theories of Magnetic Storms and Auroras

CHAPTER X

Lights from the Night Sky

CHAPTER XI

Temperature in Upper Atmosphere (Ionospheric Regions)

CHAPTER XII

Rocket Exploration of the Upper Atmosphere

CHAPTER XIII

Concluding Remarks

APPENDIX

CHAPTER I

GENERAL CONSIDERATIONS

1. INTRODUCTION—GENERAL SURVEY

The dense lower regions of the atmosphere, up to 10 to 20 km. above the surface of the earth, have long been the subject of intensive study by the meteorologists as the seat of the weather phenomena. The attenuated higher regions of the atmosphere had also been the theme of study by the geophysicists. But such studies had long been of a more or less academic character. It is only in comparatively recent years that the great importance of comprehensive study of the upper regions of the atmosphere has been recognized, because, the physical state of these regions and the phenomena occurring therein are not only of scientific interest but are also of great importance to our everyday life. To mention a few examples, long distance night radio communication is only possible because of the presence of ionized regions at and above 200 km.; auroral displays which illuminate the long winter nights in polar regions occur with greatest frequency round 100 km. and auroral streamers sometimes extend beyond 1,000 km.; shooting stars which are of such common occurrence at night appear and disappear in the region 40 to 150 km.; sound of cannon fires in France, during the Great Wars, was heard in England because the sound wave of explosion, proceeding upwards, was bent down by a hot region at 40–50 km. height; and, contemporary developments of high flying air-craft demand accurate knowledge of the physical properties of the highest regions of the atmosphere.

To the scientist the upper atmospheric regions present the picture of a vast physical laboratory where Nature is carrying out experiments on phenomena like bombardment of air masses by charged particles, electric discharge in rarefied gases, ionization by collision, magnetic double refraction and photo-chemical reactions under conditions and on a scale not attainable in an ordinary laboratory. The track of a charged particle in the Wilson cloud chamber may only be a few centimetres long; but in the upper atmosphere these may extend hundreds of kilometres. In experiments on electric discharge in the laboratory the rarefied gas has to be enclosed in a glass vessel. The walls of the vessel are then responsible for the quick disappearance of the ions and electrons and, together with it, of the discharge glow. In the rarefied upper atmosphere there are no glass walls. Hence the ions and electrons produced by any agency, continue their existence for a long time and are responsible for effects not ordinarily observable in the laboratory. In fact, the diverse problems that arise in the study of upper atmospheric phenomena are of interest to specialists in many branches of science—astrophysicists,

meteorologists, astronomers, mathematicians, geomagneticians, radio-physicists and others.

Before commencing our study it will be helpful to define the region of the atmosphere which we shall call the upper atmosphere.

It has been customary for the meteorologists, since the days of Teisserenc de Bort, to divide the atmosphere into two regions—the troposphere and the stratosphere. The division is based on the mode of temperature control in the two parts—the temperature, ultimately, being dependent on solar radiation. In the troposphere convection is in control; in the stratosphere radiation is the controlling factor.

For the purpose of the present treatise it will also be convenient to divide the atmosphere into two regions. These are to be called the lower atmosphere and the upper atmosphere. The former is identified with the troposphere and the latter will comprise the entire region lying above it. (Sometimes, the stratosphere and the region lying immediately above it will be referred to as the middle atmosphere.) This division is based on the characteristically different manners in which the solar radiation affects the two parts. In the lower atmosphere the predominant effect is heating (by an indirect process) followed by convective mixing. The solar radiation responsible for this is in the near infra-red, the visible and the near ultra-violet in the spectral range $\lambda > 3000$ Å. In the upper atmosphere there is also heating effect. But there are besides other important effects of solar radiation, e.g., photo-dissociation, allotropic modification, photo-excitation and ionization of the constituent gases. The solar radiation responsible for these is the entire ultraviolet region $\lambda < 3000$ Å. The present treatise is concerned mainly with the study of these latter effects. The stratosphere, insofar as it is concerned with the weather phenomena, will be excluded from the study.

Before closing these introductory remarks it may be useful to recall the extreme tenuity of the higher regions of the upper atmosphere. At the base of the stratosphere at 15 km. height, the density is about one-eighth of that at the ground. At 100 km. the density is one-millionth, i.e., the pressure is about 10^{-3} mm., which is of the same order as that in the so-called vacuum of ordinary electric bulb. At 300 km. the pressure is of the order 10^{-6} mm. which is that attainable in laboratory by only high quality vacuum pumps. The mean free path of a molecule in this region is thus 10^{6} cm. as compared to 10^{-5} cm. at the ground. But, as explained above, notwithstanding such high tenuity the upper atmospheric regions are the seats of many phenomena of human interest.

In Fig. 1 an attempt has been made to illustrate some of the physical features and the phenomena occurring in the upper atmospheric regions. These are, of course, illustrative and should not be taken too literally. This remark is particularly applicable to the density distribution above 100 km. which is subject to uncertainty due to uncertainty of temperature distribution and also of the mean molecular mass. While on the assumption of a diffusive equilibrium above 250 km. the region should consist

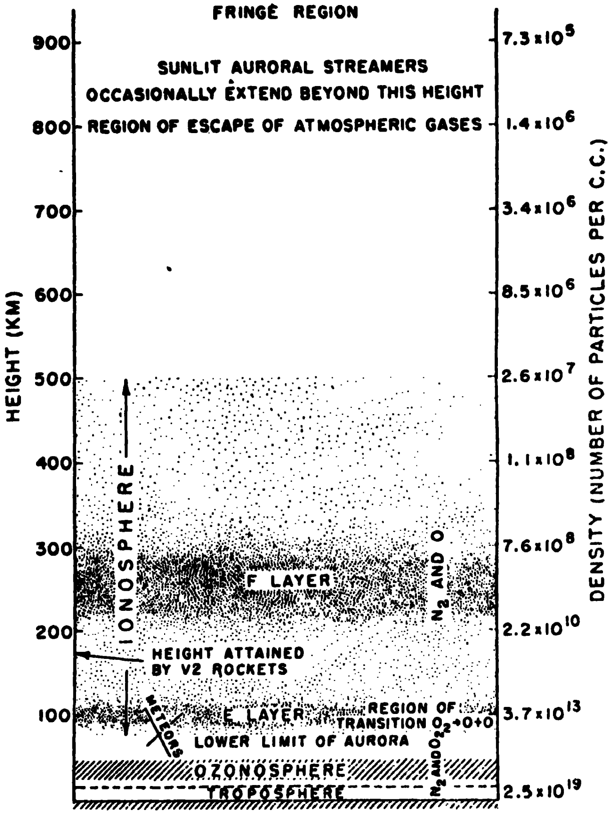


FIG. 1. Illustrating some of the physical features of the terrestrial atmosphere. Ionospheric regions are indicated by shading with dots. For the assumptions on which the molecular densities have been calculated, see Chapter XIII.

almost wholly of atomic oxygen, yet auroral spectra give unmistakable evidence of the abundance of N_2^+ ions [Chapter VIII].

2. METHODS OF INVESTIGATION

The methods of investigating the upper atmospheric regions may be classified under two heads—direct and indirect.

In the irect method the agency employed for investigation is under the control of the investigator.

In the indirect method, geophysical phenomena which are known to have some bearing on the physical state of the upper atmosphere are critically studied and the upper atmospheric conditions deduced therefrom.

A straightforward direct method is to send up high flying crafts carrying instruments for recording the atmospheric data, e.g., pressure, temperature, humidity, or any other measurable physical quantity. For this purpose the meteorologists had long been using the sounding balloons. In the modern so-called radio-sondes the balloon carries a radio transmitter which telemeters the data to the ground station. The advantage of this is that

TABLE I

Direct Methods

Region studied	Method	Information gained
0–32 km. ..	Sounding balloons.	Atmospheric constituents are uniformly mixed. Temperature falls with height (gradient—about 6°K./km.) up to *tropopause* the height of which varies from 18 km. near equator to 8–10 km. in polar regions. In the stratosphere the temperature remains nearly constant with height.
0–30 km. ..	Smoke shells.	Seasonal winds.
0–120 km. ..	High flying rockets. (Height reached 180 km.)	Composition of the atmosphere at 70 km. level is the same as in the troposphere. Temperature and density distribution up to 120 km. is of the same nature as found by indirect methods. Solar spectrum extended beyond ozone absorption limit; Mg^+ doublet (λ 2802) obtained in emission.
35–60 km. ..	Abnormal propagation of sound of explosion.	Temperature rise in the middle atmosphere; wind systems.
70–500 km. ..	Radio-wave propagation.	Atmospheric constituents are ionized; values of H (scale height), intensity of the terrestrial magnetic field, recombination coefficients of ions and electrons measured. High temperature (1000°K. *cir*) in the region 250 km. 'Bursts' of ionization are produced by meteors.

one has not to wait till the recording apparatus is brought down; instantaneous knowledge of the data is gained. The maximum height reached by such balloons (filled with hydrogen gas) seldom exceeds 32 km.

Another method which is used for study of winds up to 30 km. is to shoot up shells which burst and form a smoke cloud.

A radical advance in the method of direct study was made in 1946 when the V-2 rocket, an engine of destruction invented during World War II, was employed for upper atmospheric exploration. These rockets reach great heights (the highest record is 180 km.) and even in the very first experiments, regions up to 120 km. have been explored. A brief account of the V-2 rocket study will be given in Chapter XII.

The other agencies employed for direct study are pressure waves produced by explosion and radio wave 'pulses.' The former proceeding upwards may be bent down if it encounters a hot region of the atmosphere. From records of the downcoming wave information regarding pressure and temperature distribution in the region concerned may be deduced. The latter is now one of the most powerful direct agencies for exploring the highest regions of the atmosphere even beyond those attained by V-2 rockets. The radio wave, directed upwards, comes down if it meets

TABLE II

Indirect Methods

Region explored	Critical study of	Information gained
40–150 km. ..	Meteoric phenomena	Temperature-rise in the middle atmosphere with correspondingly higher density; temperature-drop round 80 km. level; seasonal winds.
20–60 km. ..	Spectroscopic and spectro-photometric studies of solar radiation.	Existence of ozone with maximum concentration at about 25 km. level; thickness of the ozone layer reduced to S.T.P. is only about 0·25 cm.
75–85 km. ..	Noctilucent clouds	Low temperature (200°K. *cir*); existence of high velocity winds.
50–400 km. ..	Barometric oscillations.	Tides in the upper atmosphere; rising temperature with a cold top in the middle atmosphere.
90–120 km. ..	Variations of terrestrial magnetic elements.	Electric current systems and high electrical conductivity.
60–500 km. ..	Luminescence of the night sky.	Sodium present in the region 60–120 km. Region above 100 km. contains mainly atomic oxygen and molecular nitrogen. N_2 molecules are ionized by solar rays.
80–1000 km. ..	Aurora.	Atomic oxygen, atomic nitrogen, and ionized molecular nitrogen in upper atmosphere; entry of high-speed charged particles.
800 km. (?) and beyond.	Escape of atmospheric gases; leakage of helium; zodiacal light.	The atmosphere is capped by a fringe region. High temperature (*cir* 1500°K.) near the level of escape. Zodiacal light is produced by light scattered by particles in the fringe region (?).

an ionized region of appropriate characteristics and, when it does so, brings with it indelible messages regarding physical condition of upper atmosphere which are no less useful than those recorded by instruments carried by sounding balloons or rockets.

The geophysical phenomena, the critical study of which gives valuable information regarding the state of the upper atmosphere are, meteoric flashes, aurora, luminescence of the night sky, variations of terrestrial magnetic elements, pressure oscillations of the atmosphere and absorption in solar spectrum. It is remarkable that the forms of the temperature and pressure distribution curves up to 120 km. as inferred from these indirect studies long ago, agree closely with those obtained from direct records by V-2 rockets.

The various direct and indirect methods that are employed for upper atmospheric investigation and the main results obtained therefrom are collected in Tables I and II.

3. PRESSURE AND DENSITY DISTRIBUTIONS IN SOME IDEAL CASES—THE OUTER ATMOSPHERE

(a) Some general theorems

We shall consider in this section some general theorems regarding the behaviour of a mass of gas like the atmosphere, under the permanent influence of the earth's gravitational field.

The fundamental equation connecting the pressure P and the density ρ at any height h is

$$dP = -g\rho\, dh, \qquad (1)$$

where g is gravitational acceleration.

We also have the general gas equation

$$P = nkT, \qquad (2)$$

where T—absolute temperature,

n—number of molecules per cm.^{3},

k—Boltzmann constant

$= 1{\cdot}372 \times 10^{-16}$ erg/degree.

Substituting in Eq. (1) nm for ρ where m is the mean molecular mass and dividing by (2) we have

$$\frac{dP}{P} = -\frac{mg}{kT}dh = -\frac{1}{H}\,dh, \qquad (3)$$

where H ($= kT/mg$) is the so-called *scale height* or height of the homogeneous atmosphere at temperature T. The meaning of H will be clear from the following consideration. If we assume the gas to be incompressible, then a column of gas of height H and of uniform density ρ will exert at the level under consideration the pressure P. Referred to the surface of the earth ($h = 0$) where m, the mean molecular mass of air, is $4{\cdot}8 \times 10^{-23}$ gm. and $T = 273$°K. the value of $H = 7{\cdot}9$ km. Equation (3) is very important because it provides a relationship between pressure (or density) distribution and temperature distribution with height. If one of these is determined experimentally, the other can be deduced with the help of this relation.

Let P_0, ρ_0 and T_0 be the pressure, density and temperature respectively at the surface of the earth. Then the pressure at any height h is obtained by integrating Eq. (3). Thus

$$P = P_0 \exp\left[-\frac{mg}{k}\int_0^h \frac{dh}{T}\right]. \qquad (4)$$

Also, since

$$\frac{P}{P_0} = \frac{\rho T}{\rho_0 T_0},$$

we have

$$\rho = \frac{\rho_0 T_0}{T} \exp\left[-\frac{mg}{k}\int_0^h \frac{dh}{T}\right]. \qquad (5)$$

As a special case, let the temperature increase uniformly with height. If α be the rate of increase of T with height, then $T = T_0 + \alpha h$ and therefore

$$\rho = \frac{\rho_0 T_0}{T_0+\alpha h} \exp\left[-\frac{mg}{k}\int_0^h \frac{dh}{T_0+\alpha h}\right]$$

$$= \frac{\rho_0 T_0}{T_0+\alpha h} \exp\left[-\frac{mg}{k\alpha}\log\frac{T_0+\alpha h}{T_0}\right]$$

$$= \rho_0 \left(\frac{T_0+\alpha h}{T_0}\right)^{-\left(1+\frac{mg}{k\alpha}\right)} \quad \ldots \quad (6)$$

(b) Isothermal (conductive) equilibrium

If the earth's atmosphere were at rest, uninfluenced by any external agency, then the conduction of heat from one part to another, slow as it is, would after a sufficient length of time produce uniform temperature throughout its entire mass. Further, if the atmosphere consisted of more gases than one, then the pressure and density of each gas would be distributed according to the equations,

$$\left.\begin{aligned} P^{(n)} &= P_0^{(n)} \exp\left[-\frac{m^{(n)}g}{kT}h\right] \\ \rho^{(n)} &= \rho_0^{(n)} \exp\left[-\frac{m^{(n)}g}{kT}h\right] \end{aligned}\right\} \quad \ldots \quad (7)$$

where the superscript (n) refers to the nth constituent gas of the atmosphere.

Such an atmosphere is said to be in *isothermal equilibrium* and the law of variation of density (or pressure) is known as Dalton's law.

The isothermal atmosphere, which has been assumed to start from the surface of the earth, has no natural limit. According to Eq. (7) the density falls exponentially to zero at $h = \infty$.

We note that the ratio of the densities of two gases (1) and (2) at height h is given by

$$\frac{\rho^{(1)}}{\rho^{(2)}} = \frac{\rho_0^{(1)}}{\rho_0^{(2)}} \exp\left[-(m^{(1)}-m^{(2)})\frac{gh}{kT}\right].$$

Thus, at heights sufficiently great, the lighter gas will predominate.

In Eqs. (4), (5) and (7) the value of g has been assumed to be constant with height. This is not correct. The value of g at height h is given by

$$g = g_0 \left(\frac{a}{a+h}\right)^2,$$

where a is the radius of the earth and g_0 and g refer to the values of g at earth's surface and at height h respectively; therefore we have

$$\frac{dP}{P} = \frac{d\rho}{\rho} = -\frac{m^{(n)}g_0}{kT}\left(\frac{a}{a+h}\right)^2 dh.$$

Integrating,

$$\left.\begin{aligned} P &= P_0 \exp\left[-\frac{m^{(n)}}{kT} g_0 \frac{ha}{a+h}\right] \\ \rho &= \rho_0 \exp\left[-\frac{m^{(n)}}{kT} g_0 \frac{ha}{a+h}\right] \end{aligned}\right\} . \qquad \ldots \quad \ldots \quad (8)$$

The ratio of densities of the two constituent gases at height h is now given by

$$\frac{\rho^{(1)}}{\rho^{(2)}} = \frac{\rho_0^{(1)}}{\rho_0^{(2)}} \exp\left[-\frac{m^{(1)}-m^{(2)}}{kT} g_0 \frac{ha}{a+h}\right]$$

According to Eq. (8) the density ρ even at an infinite height has a finite value. Under such condition no molecule could ever be said to have escaped from the atmosphere. It has, however, been shown by Milne [1] that if account is taken of the change in g due to change in the attraction by the mass of air itself on outer fringe as one goes outwards, then, at very great heights, the molecular density will fall off according to an inverse square law. Taking this factor into consideration, it can be shown that assuming the isothermal atmosphere to begin from the surface of the earth, a more correct expression for ρ at great heights in the outer atmosphere is given by

$$\rho = \rho_0 \left(\frac{a}{a+h}\right)^2 \exp\left[-\left(\frac{h}{a+h}\right)\left(\frac{mg_0}{kT}a-2\right)\right] . \qquad \ldots \quad (9)$$

It will be noticed that if $(a+h)$ is of the same order as a (that is, if h is within a few hundreds of kilometres), the expression for ρ reduces to the form of Eq. (8). On the other hand, when $(a+h) \gg a$,

$$\rho \doteqdot \rho_0 \left(\frac{a}{a+h}\right)^2 \exp\left[-\left(\frac{mg_0}{kT}a-2\right)\right],$$

that is, ρ varies inversely as the square of $(a+h)$.

The ratio of densities of two constituent gases at height h is given from Eq. (9), by

$$\frac{\rho^{(1)}}{\rho^{(2)}} = \frac{\rho_0^{(1)}}{\rho_0^{(2)}} \exp\left[\left(\frac{h}{a+h}\right)\left\{-\frac{(m^{(1)}-m^{(2)})}{kT} g_0 a\right\}\right].$$

The time taken for the establishment of isothermal equilibrium by diffusion of one gas through the other, each distributing itself according to Dalton's law independently of the other, will be discussed in Sec. 3.

In Eq. (1) we have neglected the centrifugal force on the molecule due to rotation of the earth. This is because even up to heights of several hundreds of kilometres from the surface of the earth this force is extremely small compared to gravity. At a sufficiently great distance from the centre of the earth however, the centrifugal force and gravity will balance one another. For a molecule lying on the equatorial plane, for instance, the balancing will occur at a distance 6·6 times the radius of the earth. Theoretically, this point will be one of minimum molecular density and beyond

this point the density will rise due to the preponderance of the centrifugal force. But we shall presently see that long before this distance is reached the atmosphere in the usual sense of the term will cease to exist due to rarity of collisions between the molecules and to their escape to the space above.

(c) Adiabatic (convective) equilibrium—the troposphere

If the atmosphere be subject to turbulence (caused by heating, etc.) and possesses convective motions, then, since conduction in gases is very slow, there is a tendency for *adiabatic equilibrium* to be set up. This is the case with the troposphere which is heated by solar radiant energy by an indirect and rather complicated process as explained below.

The major constituents of the atmosphere are extremely transparent to the visible and infra-red radiation. As such, the atmosphere (except in the very high regions) is very little heated by direct absorption. Instead, the solar radiation reaching the earth heats the ground surface and the hot surface radiates energy in the far infra-red part of the spectrum. Now, the minor constituents of the atmosphere—water vapour, carbon dioxide and ozone—have one or more strong absorption bands in this spectral region (at about $14{\cdot}7\mu$ for CO_2, $6{\cdot}5\mu$ for H_2O and $9{\cdot}7\mu$ for O_3). The outgoing radiation from the ground therefore suffers great absorption by these gases. If the temperature distribution in the lower atmosphere were determined by the emission and absorption of these long waves, then there would be, near the ground, a rapid fall of temperature with height and, at greater heights, where the absorbing gases are rare, a nearly constant temperature decreasing slowly with height. But such a rapid fall of temperature makes the distribution of density with height unstable. Convective motions are therefore set up which ultimately brings down the lapse rate to a stable value. The lower region of the atmosphere may thus be divided into two parts. In the lower part (troposphere) the temperature distribution is controlled by convection (the ground temperature being fixed by solar radiation) and in the upper part (the stratosphere) the controlling factor is radiation.

In the ideal case of adiabatic equilibrium, an element of gas, when it is transferred from one place to another, does not lose or gain any heat by conduction and takes up the requisite pressure and temperature in its new position. For an atmosphere in adiabatic equilibrium the constituent gases exist in approximately the same proportion at all heights. The case is very important since, as explained above, the temperature distribution in the troposphere tends towards this ideal condition.

We have the adiabatic relation, $PV^\gamma =$ constant. Hence

$$P = A\rho^\gamma, \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (10)$$

where,

A—a constant,

γ—ratio of the specific heat of air at constant pressure to that at constant volume.

Applying this to the general equation of equilibrium (Eq. (1)) we get

$$A\gamma\rho^{\gamma-1}\frac{\partial\rho}{\partial h} = -g\rho.$$

Integrating we get

$$\frac{A\gamma}{\gamma-1}\left(\rho^{\gamma-1}-\rho_0^{\gamma-1}\right) = -gh,$$

or,
$$\rho = \left[\rho_0^{\gamma-1}-\frac{\gamma-1}{A\gamma}gh\right]^{\frac{1}{\gamma-1}}. \quad \dots \quad \dots \quad \dots \quad (11)$$

Eq. (11) expresses the law according to which the density falls off with height in an atmosphere in adiabatic equilibrium. The constant A may be determined from the values of pressure and density at the datum level, the surface of the earth. Thus

$$A = \frac{P_0}{\rho_0^{\gamma}}.$$

In contrast to Eq. (7) we find from Eq. (11) that in this case ρ becomes zero at a finite value of h given by

$$h = \frac{A\gamma\rho_0^{\gamma-1}}{g(\gamma-1)}$$

$$= \frac{\gamma P_0}{\rho_0 g(\gamma-1)}. \quad \dots \quad \dots \quad \dots \quad (12)$$

We thus see that in the case of an atmosphere in adiabatic equilibrium we can conceive of a natural limit. If, for instance, the entire terrestrial atmosphere be assumed to be in adiabatic equilibrium, its natural limit is found from Eq. (12) to be 27·5 km. by putting $P_0 = 10^6$ dynes/cm.^{2}, $g = 981$ cm./sec.^{2} and $\rho_0 = 1{\cdot}3\times10^{-3}$ gm. per cm.^{3}

Eq. (11) also shows that for adiabatic equilibrium the temperature decreases with height. The so-called lapse rate (χ) for the ideal adiabatic atmosphere is given by

$$\chi = \frac{\partial T}{\partial h} = -\frac{g}{JC_p} = -\frac{1}{102\times10^2}\text{ degree/cm.}$$

$$= -9{\cdot}8 \text{ degrees/km.},$$

where,

J—mechanical equivalent of heat,

C_p—specific heat of air at constant pressure $= 0{\cdot}239$.

The proof of the relation is as follows:—

Combining Eq. (2) with the adiabatic equation (10), we have

$$T = \frac{Am}{k}\rho^{\gamma-1}.$$

Hence,

$$\frac{\partial T}{\partial h} = \frac{Am}{k}(\gamma-1)\rho^{\gamma-2}\frac{\partial\rho}{\partial h}. \quad \dots \quad \dots \quad \dots \quad (13)$$

Also differentiating Eq. (10) with respect to h and remembering the fundamental relation given by Eq. (1) we have

$$\frac{\partial \rho}{\partial h} = -\frac{g}{A\gamma} \cdot \frac{1}{\rho^{\gamma-2}}. \qquad \ldots \quad \ldots \quad \ldots \quad (14)$$

Substituting (14) in (13) we get

$$\frac{\partial T}{\partial h} = -\frac{mg}{k} \cdot \frac{\gamma-1}{\gamma}.$$

Now remembering that

$$\gamma = \frac{C_p}{C_v} \text{ and also } C_p - C_v = \frac{k}{Jm},$$

we have

$$\frac{\gamma-1}{\gamma} = \frac{k}{JmC_p}.$$

Thus .

$$\frac{\partial T}{\partial h} = -\frac{g}{JC_p}.$$

Actually, for the troposphere, the lapse rate is -5 degrees per km. instead of $-9{\cdot}8$ degrees per km. Further, the height of the troposphere varies between 8 and 20 km. and is not equal to 27·5 km. This shows that owing to various factors the troposphere is not in perfect adiabatic equilibrium.

We have shown above that an atmosphere in ideal adiabatic equilibrium has a natural limit. But this natural upper boundary will not be a surface separating a region of perfect vacuum above from a region containing gas molecules below. Due to thermal agitation molecules from the atmosphere below will be constantly evaporating, as it were, across the surface separating the vacuous space above from the atmosphere below in much the same manner as molecules evaporate from the body of a liquid to the space above its surface. The region immediately above the natural limit will therefore contain molecules. This region may be called the *outer atmosphere* and it will be in isothermal equilibrium. The outer layer, according to these simple considerations, needs be only a few metres thick. Account, however, has to be taken of the fact that radiation from the hot gases inside is continually coming and heating this isothermal layer and that, in order that the atmospheric arrangement may be in equilibrium, the radiation and the absorption of heat by each element must balance each other. These considerations lead to the conclusion that the isothermal layer above the adiabatic layer, instead of being a few metres thick, will extend to greater heights [6*a*].

Assuming that the atmosphere consists of two layers, the lower in adiabatic and the outer in isothermal equilibrium and working out the condition for radiation equilibrium, it has been shown that for an atmosphere of uniform constitution, the adiabatic state cannot extend to a height greater than that given by $P = \frac{1}{2}P_0$ where P_0 is the pressure at the surface of the earth. For an atmosphere of non-uniform constitution, the adiabatic layer may extend to greater heights. By assuming absorption by water

vapour of various amounts at different heights, it has been shown that the height cannot exceed that given by $P = \frac{1}{4}P_0$, which, for the case of the terrestrial atmosphere, is 10·5 km. This, as is well known, is roughly the height of the troposphere.

The outer atmosphere being more or less in isothermal equilibrium has no natural limit, though the point of minimum density due to balancing of the centrifugal and the gravitational forces is sometimes spoken of as the limit of the outer atmosphere. But as already mentioned, and as will be shown in Sec. 5, long before this height is attained (distance 6·6 earth-radii in the equatorial plane) a limit of the outer atmosphere will have been reached due to the rarity of collisions and the escape of molecules from it.

It is to be remarked in this connection that if the temperature as observed in the stratosphere remains unaltered with height at all levels, then the density in the higher regions calculated with help of Eq. (7), assumes totally unacceptable values. For example, at 300 km. level, the density becomes less than 1 gas particle per cm.^{3}. But, as we shall see in Chapter VI, there are at least 10^6 particles per cm.^{3} in this region. It therefore follows that the scale height $H(= kT/mg)$ must be greater in regions above the stratosphere. We shall see in Chapters III and XI that this, in fact, means that in regions above the stratosphere the temperature attains considerably higher values.

4. TIME FOR ATTAINING DIFFUSIVE EQUILIBRIUM

In Sec. 3, we have referred to the fact that the atmosphere left undisturbed to itself for a sufficient length of time would attain isothermal equilibrium in which each constituent gas is distributed exponentially according to Dalton's law; the lighter gases will predominate in the higher regions and the heavier ones in the lower. In this section, we shall consider the time taken for attaining such an equilibrium condition starting from the state in which the gases are mixed in the same proportion at all heights. The problem is of importance in determining the proportion in which the different constituent gases would occur in the higher regions of the atmosphere. Such determination is, however, complicated by the fact that the outer atmosphere (the stratosphere and the regions above it) is not in isothermal equilibrium. Due to diurnal heating and cooling, and tidal motions the gases in the high atmosphere—even up to the F_2 region of the ionosphere (250–400 km.)—have a tendency to be mixed up opposing the tendency to their diffusive separation by gravity. In order to estimate the proportion in which the constituent gases exist at different heights, it is necessary to know at what level tendency to diffusive equilibrium becomes important in spite of turbulence due to convective currents and tidal motions.

The problem of determining the time of attainment of diffusive equilibrium has been studied by Maris [2], by Epstein [3] and by Mitra and

Rakshit [4]. We give below the analysis and the results obtained by the last named authors.

Consider the atmosphere as a mixture of two gases labelled (1) and (2). At a height h above some datum level, let $(N_{1h})_{\underset{D}{\uparrow}}$ be the number of molecules of gas (1) diffusing upwards per unit time across unit area of the horizontal plane. Then,

$$(N_{1h})_{\underset{D}{\uparrow}} = -D\,\frac{\partial n_{1h}}{\partial h}, \qquad \ldots \qquad (15)$$

where n_{1h} is the molecular density of gas (1) at the height h and D is the coefficient of diffusion (Meyer's formula, corrected), given by

$$D = \frac{1{\cdot}34}{3}\,\frac{n_{1h}\lambda_{2h}\bar{v}_2 + n_{2h}\lambda_{1h}\bar{v}_1}{n_{1h}+n_{2h}} \qquad \ldots \qquad (16)$$

where n_{2h} is the molecular density of gas (2) at height h; λ_{1h}, λ_{2h} the mean free paths of the two gases and $\bar{v}_1$, $\bar{v}_2$ their mean molecular velocities. Since $\partial n_{1h}/\partial h$ is negative, $-D\,\partial n_{1h}/\partial h$ is a positive quantity.

The molecules of gas (1) will also be falling downwards due to gravity. Let $(N_{1h})_{\underset{G}{\downarrow}}$ be the number of molecules falling down across unit area in unit time. Then

$$(N_{1h})_{\underset{G}{\downarrow}} = v_1 n_{1h}, \qquad \ldots \qquad (17)$$

where v_1 is the effective velocity of the fall. The net number of molecules going up across unit area in unit time is therefore,

$$(N_{1h})_{\uparrow} = -D\,\frac{\partial n_{1h}}{\partial h} - v_1 n_{1h}. \qquad \ldots \qquad (18)$$

The form of the expression for v_1 can be obtained from the equilibrium condition when $N_{1h} = 0$. Now, in the case of linear rise of temperature n_{1h} and n_{2h} are given by Eq. (6),

$$\left.\begin{aligned} n_{1h} &= n_{10}\left(\frac{T}{T_0}\right)^{-\left(1+\frac{m_1 g}{k\alpha}\right)} \\ n_{2h} &= n_{20}\left(\frac{T}{T_0}\right)^{-\left(1+\frac{m_2 g}{k\alpha}\right)} \end{aligned}\right\}, \qquad \ldots \qquad (19)$$

where m_1 and m_2 are the molecular masses of the constituents (1) and (2) and n_{10} and n_{20} their densities at the datum level.

Substituting in Eq. (18) the values of n_{1h} and $\partial n_{1h}/\partial h$ obtained from Eq. (19) we get, for equilibrium condition,

$$0 = D\,\frac{m_1 g + k\alpha}{kT}\,n_{1h} - v_1 n_{1h}$$

or,

$$v_1 = D\,\frac{m_1 g + k\alpha}{kT}.$$

TABLE I

Direct Methods

Region studied	Method	Information gained
0–32 km. ..	Sounding balloons.	Atmospheric constituents are uniformly mixed. Temperature falls with height (gradient—about 6°K./km.) up to *tropopause* the height of which varies from 18 km. near equator to 8–10 km. in polar regions. In the stratosphere the temperature remains nearly constant with height.
0–30 km. ..	Smoke shells.	Seasonal winds.
0–120 km. ..	High flying rockets. (Height reached 180 km.)	Composition of the atmosphere at 70 km. level is the same as in the troposphere. Temperature and density distribution up to 120 km. is of the same nature as found by indirect methods. Solar spectrum extended beyond ozone absorption limit; Mg^+ doublet (λ 2802) obtained in emission.
35–60 km. ..	Abnormal propagation of sound of explosion.	Temperature rise in the middle atmosphere; wind systems.
70–500 km. ..	Radio-wave propagation.	Atmospheric constituents are ionized; values of H (scale height), intensity of the terrestrial magnetic field, recombination coefficients of ions and electrons measured. High temperature (1000°K. *cir*) in the region 250 km. 'Bursts' of ionization are produced by meteors.

one has not to wait till the recording apparatus is brought down; instantaneous knowledge of the data is gained. The maximum height reached by such balloons (filled with hydrogen gas) seldom exceeds 32 km.

Another method which is used for study of winds up to 30 km. is to shoot up shells which burst and form a smoke cloud.

A radical advance in the method of direct study was made in 1946 when the V-2 rocket, an engine of destruction invented during World War II, was employed for upper atmospheric exploration. These rockets reach great heights (the highest record is 180 km.) and even in the very first experiments, regions up to 120 km. have been explored. A brief account of the V-2 rocket study will be given in Chapter XII.

The other agencies employed for direct study are pressure waves produced by explosion and radio wave 'pulses.' The former proceeding upwards may be bent down if it encounters a hot region of the atmosphere. From records of the downcoming wave information regarding pressure and temperature distribution in the region concerned may be deduced. The latter is now one of the most powerful direct agencies for exploring the highest regions of the atmosphere even beyond those attained by V-2 rockets. The radio wave, directed upwards, comes down if it meets

diffusion. Now, the total number of molecules of gas (1) to be separated for complete diffusion is given by

$$\left[(p_{1h})_{\eta=0}-(p_{1h})_{\eta=1}\right]\times \text{number of molecules of gas (1) to produce unit pressure},$$

where p_{1h} is the partial pressure of gas (1) at level h, and hence the time for the first 10 per cent diffusion is easily calculated. Similarly, the times for the successive 10 per cent diffusions are found out. The total time required for 80 per cent diffusion is then calculated for different levels.

It should be remembered that with the progress of diffusive separation, the temperature distribution tends to become constant with height. This has not been considered in the calculation. Since with the progress of diffusion the upper part becomes cooler and the rate of diffusion becomes slower, the time for diffusive separation would, in general, be greater than that given in Table III.

In order to carry out numerical calculations it is necessary to know the physical conditions, viz., the composition, pressure and temperature at the datum level and also the probable distribution of temperature above this level. The assumptions made by Mitra and Rakshit and the reasons for making the same are given below briefly. For fuller details the original paper may be consulted.

(1) The 100 km. level is taken as the datum level.

It is assumed that the atmosphere below this level is thoroughly mixed while that above is in diffusive equilibrium.

The reason for making this assumption is that the pressure in the regions below is such that the constituents if once mixed up by turbulence will take a long time—measured in years—to settle down to perceptible diffusive equilibrium. And, observations on drifts of noctilucent clouds and meteor trails show that the regions immediately below 100 km. are subject to strong winds. The upper limit of the level up to which the atmospheric constituents should remain mixed in more or less the same proportion as at the surface of the earth cannot therefore be much below 100 km.

(2) The atmospheric constituents below 100 km. are N_2 and O_2 while those above are N_2 and O [see Chapter V]. There is, of course, a region of transition in which preponderance of O_2 gradually gives way, from bottom to top, to preponderance of O. The effect of this region of transition is neglected in the calculation.

For the sake of comparison calculations are also given for the case of an atmosphere consisting of N_2 and O_2.

(3) The pressure and temperature at 100 km. are taken as 10^{-3} mm. and 300°K. respectively [see Chap. XIII, Sec. 1].

(4) The numbers of N_2 molecules and O atoms at the datum level are taken as $2{\cdot}54\times10^{13}$ and $6{\cdot}96\times10^{12}$ per cm.^{3} respectively.

(5) The temperature is assumed to increase linearly with height with a gradient of 4°K. per km. from 100 km. [see Chap. XIII].

Tables III and IV give the results obtained by Mitra and Rakshit.

TABLE III

Time for 80% *diffusion*

Height (km.)	N_2, O Rising temperature	N_2, O_2 Rising temperature	N_2, O_2 Constant temperature
165	....	....	14·12 hours
175	....	68·26 hours	5·53 "
200	....	....	0·46 "
225	....	10·09 hours	....
250	17·47 hours	3·90 "	....
300	7·98 "	....	....
350	3·67 "	....	....

TABLE IV

Percentage diffusion at different levels after 10 *hours*
(*Rising temperature*)

Height (km.)	N_2, O	N_2, O_2
100	0	0
125	3·5%	2·5%
150	10·5%	9%
200	43%	47%
250	71%	95%
300	82%	Practically complete diffusive equilibrium.
350	91%	
400	....	

If we assume that the upper atmosphere is disturbed by periodic forces (due to diurnal heating and cooling and tidal motions) and that as a consequence the constituents are mixed, then on the assumption of a rising temperature, the constituents have no time to settle by diffusive separation even up to the height of 250 km. Above this height, the time required for diffusive separation becomes small compared to the period of the disturbing force. Diffusive separation may therefore become important beyond this height. On the basis of the above considerations one may expect that N_2 and O above 250 km. are more or less in diffusive equilibrium and that in the highest regions of atmosphere atomic oxygen will predominate.

5. ESCAPE OF GASES EROM TERRESTRIAL ATMOSPHERE—THE FRINGE OF THE ATMOSPHERE

(*a*) Level of escape of atmospheric gases

A question of considerable interest and importance is the height of the so-called *ceiling* or the *top* of the atmosphere, i.e., the height above which the atmosphere in the ordinary sense of the term ceases to exist and at which its

limit may be said to have been reached. An atmosphere in ideal adiabatic equilibrium has got a natural limit; this, as we have seen, is obtained from Eq. (11) by putting the density ρ equal to zero. The outer atmosphere being in more or less isothermal equilibrium has, however, no natural limit, i.e., no finite height at which ρ becomes equal to zero (Eq. (7)). We may, however, imagine a limit from the following considerations. In any region above the surface of the earth, the atmosphere in the ordinary sense of the term can exist only if the molecules in the region are prevented from escaping by collisions with the molecules above. Now, as one goes up, the atmosphere becomes increasingly rarefied and the frequency of collisions between the gas particles becomes smaller and smaller. A height is ultimately reached where the collisions are so few and far between that a molecule moving upwards, by receiving an impact from a molecule in the denser atmosphere below, has little chance of returning to earth by collision with molecules above. The height at which this state of affairs prevails may be said to be the limit of the atmosphere. Above this height the molecules move freely with the velocity acquired at the last collision in the lower region and, being subject only to the pull of gravity, describe parabolic, elliptic or hyperbolic paths according to the magnitude of their velocities. This region is called the *fringe* region of the atmosphere or the *exosphere*. Molecules with hyperbolic orbits will, of course, escape from the earth and the velocity necessary for escape is given by

$$v^2 > \frac{2ga^2}{r}$$

where v—velocity of the molecule,
g—acceleration due to gravity at the level from which the molecule escapes,
a—earth's radius,
r—distance of the level of escape from the centre of the earth.

The minimum velocity v_c for escape—the so-called *critical* velocity is thus about 11 km./sec. (r = 7,378 km., a = 6,378 km., g = 733 cm./sec.^{2} at 1,000 km. level).

Cone of escape

The first attempt to calculate the height at which collision begins to have negligible effect was made by Milne [1] and was later followed up by J. E. Jones [5]. It was shown by the latter that the failure of previous workers to estimate the height of the ceiling of the atmosphere or the level above which there is very little chance of a collision is due to their assumption of uniform molecular density along the free path of a molecule. At high altitudes of the atmosphere, the free path of a molecule is large and the molecular density also decreases rapidly upwards. Along the free path of a molecule, therefore, the molecular density is liable to change appreciably. Account has thus to be taken of the fact that the probability of collision of a molecule is a function not only of its velocity but also of its origin and

direction of motion, since the molecular density will vary in different manners in different directions. These considerations lead to the idea of the *cone of escape* which may be understood as follows. Let us suppose that an observer ascends upwards from a layer where molecular density is small but appreciable. If the molecules were opaque, then the hemispherical sky above the observer would appear to him to be absolutely opaque; a line drawn from the observer towards any direction would pass through many molecules, one behind the other. If the observer continues his ascent, the molecules overhead will gradually thin away and he will reach a level where the molecules overhead will just fill the sky, i.e., a line drawn upwards will pass through only one molecule, whereas a line drawn in any other direction will still pass through more than one molecule. Mounting still higher, he will find his sky overhead gradually clearing up and he will 'see' a cone with its axis vertical within which his sky will be clear. It is obvious that this cone will open out with height and the observer will finally have his whole sky clear. This is the *cone of escape* of the molecules. A molecule moving within the solid angle of this cone will have some chance of escaping without a collision.

Jones starts with the well-known expression of Tait giving the probability of collision of a molecule of velocity v with any other molecule in the interval of time δt. We have

$$f(v) = n\sigma^2\sqrt{\frac{m^3}{8\pi k^3T^3}} \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} \int_{v'=0}^{\infty} \exp\,(-mv'^2/2kT)\; v'^2 V\, dv' \sin\theta\, d\theta\, d\phi,$$

where $f(v)\delta t$ is the chance of a collision of a molecule of velocity v with any other molecule in the time δt,

n—molecular density,

σ—molecular diameter,

m—molecular mass,

v, v'—velocities of the colliding molecules,

V—relative velocity between the colliding molecules given by $V^2 = v^2+v'^2-2vv' \cos\theta$,

θ, ϕ—Eulerian angular co-ordinates defining the direction of motion of v' with respect to v.

The expression for $f(v)$ on integration reduces to the form,

$$f(v)=\frac{\sqrt{\pi}n\sigma^2}{\frac{vm}{2kT}}\Phi\left(v\sqrt{\frac{m}{2kT}}\right),$$

where Φ is of the form,

$$\Phi(x)=x\exp\,(-x^2)+(2x^2+1)\int_0^x \exp\,(-y^2)\,dy.$$

(The derivation of the expression, as also values of $\Phi(x)$ and $x^2/\Phi(x)$ are to be found in Jeans' *Dynamical Theory of Gases*, 3rd Edition, p. 255.)

2B

Since the chance of collision of a molecule of velocity v with any other molecule in traversing the path element δs is $f(v)\ \delta s/v$, the collision is a certainty within the free path s when

$$\int_0^s \frac{f(v)}{v}\, ds = 1.$$

For the case of uniform molecular density (n constant with respect to the path s) and no external field of force, the length of this path $\lambda_v \left(= \int_0^s ds\right)$ is obviously equal to $v/f(v)$.

If n is not constant, as is the case in the upper atmosphere, the length of the path is obtained from

$$\int_0^s n(s)ds = \frac{\dfrac{v^2 m}{2kT}}{\sqrt{\pi}\sigma^2 \Phi\left(v\sqrt{\dfrac{m}{2kT}}\right)},$$

where $n(s)$ denotes molecular density as a function of the path s.

Now, for variation of n with height we have Milne's expression [1], deduced by taking into account the change in g caused by the change in the attraction by the mass of air left behind as one goes outwards (Eq. (9)),

$$n = n_0 \left(\frac{r_0}{r}\right)^2 \exp\left[-\left(1-\frac{r_0}{r}\right)\left(\frac{mg_0}{kT}r_0 - 2\right)\right]$$

where

n_0—molecular density at the base of the isothermal part of the atmosphere at distance r_0 from the centre of the earth,

n—molecular density at a distance r where collision is being considered;

or, putting

$$\frac{mg_0}{kT}r_0 - 2 = q'_0,$$

$$n = n_0\left(\frac{r_0}{r}\right)^2 \exp\left[-\left(1-\frac{r_0}{r}\right)q'_0\right] \quad \cdots \quad \cdots \quad (21)$$

Now, consider a molecule which has acquired its velocity by collision at a point distant r from the centre of the earth, and moving in a direction making angle θ with the radial direction at the point. If the path s is reckoned from the point of collision and R denotes the radial distance corresponding to s, then, neglecting the curvature of layers of equal density,

$$R = r + s\cos\theta.$$

The length of the free path is then given by

$$\int_0^s n(R)ds = \frac{\dfrac{v^2 m}{2kT}}{\sqrt{\pi}\sigma^2 \Phi\left(v\sqrt{\dfrac{m}{2kT}}\right)},$$

where $n(R)$ denotes the molecular density at a distance R.

Introducing the expression for n as given in Eq. (21) and integrating, we obtain

$$\exp\left(q'_0 r_0/r\right) - \exp\left(q'_0 r_0/R\right) = \frac{q'_0 e^{q'_0} \cos\theta \dfrac{v^2 m}{2kT}}{\sqrt{\pi} n_0 \sigma^2 r_0 \Phi\left(v\sqrt{\dfrac{m}{2kT}}\right)}. \quad .. \quad (22)$$

This equation determines R and therefore the free path s from the relation

$$s = (R - r) \sec\theta,$$

when r, v and θ are given.

If the molecule is to escape from the atmosphere, then s and, therefore, R should be infinite. We thus get from Eq. (22), the condition for the escape of the molecule,

$$\exp\left(q'_0 r_0/r\right) - 1 = \frac{q'_0 e^{q'_0} \cos\theta \dfrac{v^2 m}{2kT}}{\sqrt{\pi} n_0 \sigma^2 r_0 \Phi\left(v\sqrt{\dfrac{m}{2kT}}\right)}. \quad .. \quad (23)$$

The estimate of the lowest level from which escape is possible is obtained by putting $\theta = 0$ and $v = \infty$ in the above equation, i.e., when the molecule is moving radially with infinite velocity. Since for $x = \infty$, $x^2/\Phi(x)$ is $\pi^{-\frac{1}{2}}$ (Jeans, *loc. cit.*), we have from Eq. (23) for this case,

$$\exp\left(q'_0 r_0/r_c\right) - 1 = \frac{q'_0 \exp\left(q'_0\right)}{\pi n_0 \sigma^2 r_0}, \quad .. \quad .. \quad (24)$$

where r_c is the height of the lowest or the *critical level* of escape.

Introducing in the relation the expression for the variation of molecular density with height as given by Eq. (21), we have for the density n_c at the critical level,

$$n_c = \left(\frac{r_0}{r_c}\right)^2 \left\{n_0 e^{-q'_0} + \frac{q'_0}{\pi\sigma^2 r_0}\right\}.$$

Hence, if q'_0 is large, the molecular density n_c at the critical level is given by

$$n_c \doteq \frac{q'_0}{\pi\sigma^2 r_0}.$$

It is interesting to note that in such case, since the order of the expression is determined almost entirely by σ^2 and by r_0 which are approximately the same for all molecules, the density at the critical level is the same for all kinds of molecules, namely, of the order of 10^8 per c.c.

At the *critical height* r_c, a molecule in order to escape must move radially ($\theta = 0$) and with infinite velocity. Above the critical height there is a range of values of θ for which escape is possible. The limiting value of θ, for r greater than r_c is obtained from Eq. (23) by putting $v = \infty$. Thus

$$\cos\theta_r = \frac{\pi n_0 \sigma^2 r_0}{q'_0 \exp\left(q'_0\right)} \left[\exp\left(q'_0 r_0/r\right) - 1\right].$$

Or, introducing the critical height r_c from Eq. (24)

$$\cos \theta_r = \frac{\exp(q'_0 r_0/r)-1}{\exp(q'_0 r_0/r_c)-1}. \qquad \ldots \quad \ldots \quad (25)$$

The cone of semi-angle θ_r given by the above equation is the *cone of escape*.

Thus, in order that a molecule *may* escape from the atmosphere,

(i) its velocity must exceed the critical value $\left(v^2 > \frac{2ga^2}{r_c}\right)$,

(ii) its last point of collision must be above the critical level r_c (Eq. (24)), and,

(iii) its direction of motion, after last collision, must fall within the cone of escape (Eq. (25)).

It should, however, be noted that not *every* molecule satisfying the above conditions will be able to escape. The angle θ_r of the cone (Eq. (25)) is the maximum that is *possible* for escape (when the velocity is maximum, i.e., $v = \infty$). For a molecule with finite velocity v_θ, the direction must fall within a cone of smaller angle ($\theta < \theta_r$), in order that its subsequent motion may be free from collision. The value of this angle is easily obtained from Eq. (23).

In the discussions above we have called the lowest level of the region of escape (which may be of considerable thickness) as the critical level. In discussions on the escape of planetary atmosphere, however, the middle, instead of the lowest level of this region, is often referred to as the critical level. In the discussions that follow we will refer to this middle level as the mean critical level.

Jones [5] and Milne [1] calculated the levels of escape of hydrogen and helium, which at the time they wrote their papers were believed to be the probable constituents of the upper atmosphere. Assuming the temperature of the outer atmosphere to be 219°K they obtained the value of the level of escape to be 1,521 km. if the outer atmosphere consisted of hydrogen, and 630 km. if it consisted of helium. The heights are measured from above the stratosphere, 20 km. from the surface of the earth. The assumptions of Milne and Jones regarding the constituents and temperature distribution in the upper atmosphere are inconsistent with later observations. We know now that the atmosphere above 100 km. consists of molecular nitrogen, atomic oxygen and also probably atomic nitrogen. The gases are more or less in diffusive equilibrium and are at a high temperature of the order of 1000°K. Ionospheric observations give reliable estimate of the molecular density at the 100 km. level (about 10^{13} per cm.^{3}). If this level is taken as the datum level then the disposition of the gases with height can be calculated for any probable rise of temperature with height. The minimum height of the level of escape of atomic oxygen, which is the predominant gas in

the outer region of the upper atmosphere, is thus found to be 671 km. above the datum level at 100 km. ($\sigma = 2{\cdot}63\times10^{-8}$ cm., $T = 1000°$K).

(b) The problem of helium

It is obvious that molecules above the critical height moving with velocity greater than the critical velocity within the cone of escape will escape from the earth. And, since, whatever be the mean velocity of the gas molecules (i.e., the temperature), a certain fraction of the molecules will always have velocities greater than the critical velocity, this fraction will always be escaping from the earth. The important question—how long will it take for a gas with a certain velocity to be dissipated from the earth?—therefore, arises in this connection.

The question is of particular importance to light gases like hydrogen and helium, which on account of their higher mean velocity have greater chance of escape.

The problem was first attacked by Jeans [6] in a simple manner and later by Milne [1] and by Jones [5] taking into account the cone of escape discussed in the previous section. The time required for the dissipation of a gas as obtained by the more rigorous calculations of Milne and of Jones is in all cases longer than that obtained by Jeans. For instance, for the escape of hydrogen assuming the upper atmosphere to be at a temperature of $-54°$C., Jones finds that the concentration would fall from $1{\cdot}89\times10^{13}$ to $1{\cdot}89\times10^{8}$ in 2×10^{24} years; Jeans, on the other hand, finds that for total dissipation $2{\cdot}8\times10^{24}$ years will be required. However, notwithstanding the greater rigour of Milne and Jones' method, we briefly describe below the method of Jeans' calculation because it is based on very simple ideas and will give results sufficiently approximate for our purpose.

The number of molecules which cross unit area of the sphere of radius r_c per unit time with velocity greater than the critical velocity is given by

$$n\left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}}\int\int\int e^{-\frac{m}{2kT}(v_x^2+v_y^2+v_z^2)}\,v_z dv_x dv_y dv_z$$

where,

n—molecular density at the sphere $r = r_c$,

v_x, v_y, v_z (z direction along the radius of the sphere)—components of velocity of the molecule along the co-ordinate axes, and

$$v_x^2+v_y^2+v_z^2 > \frac{2ga^2}{r_c}.$$

This number of molecules, according to Jeans, is escaping from the earth's atmosphere per unit area per unit time.

Changing into spherical co-ordinates and integrating, this is given by

$$\frac{n}{2\sqrt{\frac{\pi m}{2kT}}}\,e^{-\frac{mga^2}{kTr_c}}\left(1+\frac{mga^2}{kTr_c}\right).$$

Substituting for n from Eq. (8), namely,

$$n = n_0 \exp\left(-\frac{mga}{kT}\,\frac{r_c-a}{r_c}\right)$$

where n_0—molecular density at the base of the isothermal atmosphere, we have the number of molecules that is lost per unit time per unit area from the atmosphere

$$\frac{n_0}{2\sqrt{\frac{\pi m}{2kT}}}\,e^{-\frac{mga}{kT}}\left(1+\frac{mga^2}{kTr_c}\right). \quad \ldots \quad \ldots \quad (26)$$

Therefore the time t_0 required to lose the molecules that are contained in a layer of unit thickness at the base of the isothermal atmosphere is

$$t_0 = \frac{n_0}{\dfrac{n_0}{2\sqrt{\frac{\pi m}{2kT}}}\,e^{-\frac{mga}{kT}}\left(1+\dfrac{mga^2}{kTr_c}\right)} = \frac{2\sqrt{\frac{\pi m}{2kT}}}{1+\dfrac{mga^2}{kTr_c}}\,e^{\frac{mga}{kT}} \text{ sec.}$$

Or, substituting for T the value $mC^2/3k$ where C is the root mean square velocity of molecules at temperature T and remembering that a/r_c is nearly equal to unity,

$$t_0 = \frac{4.34}{C\left(1+\dfrac{3ga}{C^2}\right)} \exp\left(3ga/C^2\right) \text{ sec.} \quad \ldots \quad \ldots \quad (27)$$

If H be the height of the homogeneous atmosphere, the time t_1 required for the whole atmosphere to escape is equal to $t_0 \times H$.

From the expression for t_0 it is easily found that if the mean velocity is equal to the critical velocity then the gas will be entirely dissipated in 1·4 hours. The time of dissipation, however, rapidly increases as the mean velocity becomes smaller. If, for instance, the mean velocity be one-fifth that of the critical velocity then it will take $1{\cdot}9\times10^{10}$ years for the gas to escape. We shall now use the above analysis to study the case of the escape of helium from terrestrial atmosphere.

Careful measurements show that the helium content near the surface of the earth is 5×10^{-4} per cent by volume. This proportion of helium is found to exist at least up to about 70 km. as measurements on samples of

air collected in rocket flights show [6b]. If the height of the homogeneous atmosphere is taken as 7·9 km., the amount of helium per sq. cm. column on the surface of the earth is found to be 10^{20} atoms. This is considerably less than the amount that must have been discharged in geological times from the crust of the earth by the washing away of the igneous rocks which contain uranium and thorium. According to an estimate by Lindemann [7], the helium content of the atmosphere due to this discharge ought to have been at least 6×10^{21} helium atoms per sq. cm. column. To this should be added the helium which is being constantly discharged with the natural gases from the so-called helium fields of North America. This amounts to 2×10^{7} cu. m. per annum. But according to Lindemann (*loc. cit.*) it is unlikely that this discharge could have affected the total averaged over geological periods—10^{9} years. Since the total content of helium in the atmosphere at a very liberal estimate is 5×10^{20} atoms per sq. cm. column, it follows that at least $5{\cdot}5 \times 10^{21}$ helium atoms must have disappeared from the terrestrial atmosphere. This means that from the outermost layers 10^{5} atoms must have been disappearing per sq. cm. per sec.

The two obvious modes of disappearance of helium are its chemical combination with other atmospheric gases and its escape from the terrestrial atmosphere. The first of these possibilities is ruled out because of the chemical inertness of the gas. Regarding the second, from the expression (26) we find that helium can escape at this rate from the earth if the temperature of the escaping helium be greater than 1000°K. The problem of escape has also been examined critically by Spitzer [7a] after modifying the treatment given above. According to Spitzer the temperature at the critical level of escape is of the order 1500°K. We shall see in Chapter XI that a number of other atmospheric phenomena also points to the existence of a high temperature of this order in the upper regions.

Helge-Peterson [8] has closely examined the case of the escape of helium by steady upward diffusion through the rest of the atmospheric gases. He comes to the remarkable conclusion that in such case, when a steady state has been reached, the decrease of helium density with height will be much more rapid than that for the normal case of no escape with the helium in static equilibrium according to its own partial pressure. In fact, for the case of escape, the density of helium would be 10^{6} to 10^{10} times smaller than the case of no escape with static equilibrium. This conclusion, which is by no means obvious without detailed calculation, explains the scarcity of helium in the upper regions of the atmosphere as indicated by the absence of helium lines in the spectrum of the night sky and of aurora.

The problem of the escape of atmospheric gases has been re-examined in recent years by Spitzer [7a]. According to Spitzer the previously calculated values of the time of escape (t_1) are too low and have to be multiplied by a factor 10^{6}. This is because while the upper regions of the atmosphere from where escape takes place is isothermal at a high temperature, the underlying region is at a much lower temperature. The

density at these lower heights is thus much greater than that of the hypothetical atmosphere assumed by previous workers and the time necessary for escape is thus also increased.

With the above correction a temperature of 1500°K is needed at the critical height for escape for helium. Spitzer, however, estimates that the average temperature in the region of escape may not be much above 500°K. During solar flares when intense ultra-violet light is emitted the temperature at and above 300 km. is raised to 2000°K and helium escapes during these periods of ultraviolet blasts.

(*c*) The 'fringe' region or the exosphere

It is evident that molecules or atoms moving with velocity less than the critical velocity may also avoid collision if they move within the cone as defined by Eq. (26). Such particles will not, however, be able to escape from the atmosphere but will, after describing elliptic paths, fall back to the atmosphere (Fig. 2). These high flying particles constitute what may be called the *fringe* of the atmosphere or the *exosphere*. The fringe region obviously commences from the level where the semi-vertical angle of the cone of escape approaches 90°. The heights to which the particles will rise will depend on the magnitude and the direction of velocity acquired at the last collision and also on the value of g at the point of collision.

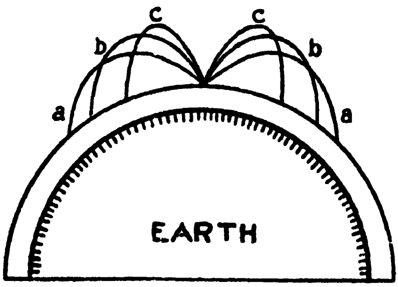


FIG. 2. Trajectories of a particle projected with velocity 3·5 km./sec. from the level of escape assumed to be 1,000 km. above the surface of the earth. The angles to the vertical are (*a*) 60°, (*b*) 45° and (*c*) 30°.

The particles in the fringe region having little chance of collision will not be subject to the gas kinetic laws. Their average density will diminish as one moves up from the region of escape. It is possible to make an estimate of the rate of decrease of the average density by making plausible assumptions regarding the density, the temperature and the velocity distribution of the particles in the region of escape which is the source of the particles in the spray region. Table V prepared by Mitra and Banerjee [9] shows how the average density of the particles in the successive layers of the spray region may be expected to diminish as one ascends into the region. The following assumptions have been made in the preparation of the table:

Source of the particles in the fringe region ..	Top layer of the region of escape
Height of the source	800–1000 km.

Average density (the particles are mostly O-atoms) in the source	3×10^5 per cm.^{3}
Temperature of the source	1000°K
Velocity distribution of the particles in the source	Maxwellian

TABLE V

Region above earth's surface in km.	v_1	v_2	n_1	n_2	Average density (number of particles per cm.^{3}).
1,250–1,000	1·50	1.87	10^{13}	$7{\cdot}8 \times 10^{12}$	$1{\cdot}0 \times 10^5$
1,500–1,250	1·87	2·60	$7{\cdot}8 \times 10^9$	$4{\cdot}0 \times 10^9$	$3{\cdot}0 \times 10^4$
1,750–1,500	2·60	3·15	$4{\cdot}0 \times 10^9$	$2{\cdot}4 \times 10^8$	$8{\cdot}0 \times 10^3$
2,000–1,750	3·15	3·59	$2{\cdot}4 \times 10^8$	$1{\cdot}4 \times 10^7$	$4{\cdot}5 \times 10^2$
2,250–2,000	3·59	3·95	$1{\cdot}4 \times 10^7$	$1{\cdot}1 \times 10^6$	$2{\cdot}3 \times 10$
2,500–2,250	3·95	4·27	$1{\cdot}1 \times 10^6$	$9{\cdot}2 \times 10^4$	2·0
2,750–2,500	4·27	4·54	$9{\cdot}2 \times 10^4$	$9{\cdot}9 \times 10^2$	$1{\cdot}0 \times 10^{-1}$

The last column in the table gives the average density of the particles in the regions corresponding to those shown in the first column. v_1 and v_2 give the range within which the velocity of the particles must lie in order that the top part of their trajectories will lie within the region concerned. n_1 and n_2 give the numbers of particles with velocities v_1 and v_2 respectively, at the 1,000 km. level, crossing over to the fringe region per sq. cm. per second.

It will be seen from the table that the density in the fringe region diminishes with increasing height and in the region of 2,500 km. from the earth's surface, the density falls to about one particle per cm.^{3} which is the order of the density of matter in the interstellar space [10].

It is to be emphasized that the values given in Table V are only illustrative and are not to be taken too literally. They depend upon the assumptions made regarding the critical height (height of the level of escape) and on the density and the temperature of the particles there. All these values are still far from certain. For example, Spitzer [7*a*] from various considerations comes to the conclusion that the mean critical level lies well above 300 km., somewhere between 500 and 1000 km. The density of the particles at this level, he assumes to be 3×10^7/cm.^{3} If these data are used the values in Table V above have necessarily to be modified.

Spitzer has also drawn attention to the shortness of the average time which the particles spend at great heights. This time, above a certain level, depends upon the density of particles at that level. According to Spitzer's computation, the time spent above the mean critical level (where density is 3×10^7/cm.^{3}) is a few minutes; above the region where density is 10^{10}/cm.^{3}, it is a few days and above the region where the density is 10^{12}/cm.^{3} about a year. According to Spitzer, the height where the

density is about 10^9/cm.^{3}, is probably between 300 and 400 km. Diffusive equilibrium is appreciable above this level, though the lower level of the same may be 150 km.

It has also been pointed out by Spitzer that the distribution of neutral atoms in the exosphere above the critical level is that of an isothermal gas with the temperature the same as about the critical level. This follows from Liouville's theorem, according to which 'the density of particles in phase space is constant along a dynamical trajectory. Thus, the density of particles moving in a certain direction at a certain height in the exosphere is exactly the same as the corresponding density for these same particles when they first entered the exosphere'. Hence if at the critical level the velocity was Maxwellian, it will also be so in the exosphere.

The region between 300 km. level and the mean critical level (lying between 500 and 1000 km. according to Spitzer) is the transition region between the normal lower atmosphere and the exosphere. In view of the small average time which the atmospheric particles spend in this region, dissociation and recombination are improbable. Hence the conditions at these heights are determined by the condition of the atmosphere below rather than by local equilibrium.

(d) Viscous drag due to rotation of the earth

An interesting point to be considered is the possibility of different layers of the upper atmosphere at different heights moving with different angular velocities on account of insufficient viscous drag. The expression for the coefficient of viscosity of a gas being $\eta = \frac{1}{3}nm\bar{v}\lambda$, it would appear that the viscosity is independent of pressure; here $\bar{v}$ is the mean velocity of the molecules and λ the mean free path. The above expression will, of course, hold as long as the gas obeys gas kinetic laws, i.e. as long as collisional frequency is appreciable. It may be noted here that in experimental investigations on the viscosity of a gas with laboratory apparatus, the coefficient of viscosity falls to a low value at low pressure. It may be shown, however, that this is due to the size of the vessel becoming comparable with the mean free path of the gas molecules. As long as the size of the vessel is large compared with the mean free path (as is the case for the atmosphere covering the earth), the coefficient of viscosity should remain independent of pressure. The effective drag of one layer of gas upon another depends not on the coefficient of viscosity, but on the so-called coefficient of *kinematical viscosity* [11] which is defined as η/ρ. It would thus appear that the drag is actually greater in the high atmosphere than in the lower atmosphere. The curve in Fig. 3 shows the change of kinematical viscosity with height for the assumed distribution of atmospheric pressure.

The assumption is therefore justifiable that up to the limit of the outer atmosphere within which collision is appreciable, the air molecules participate in the rotation of the earth. A molecule moving in the fringe region with the velocity acquired at the last collision will thus have a component velocity

in the direction of rotation of the earth. As a consequence, the trajectory of a molecule shot up radially into the fringe region will have a curvature

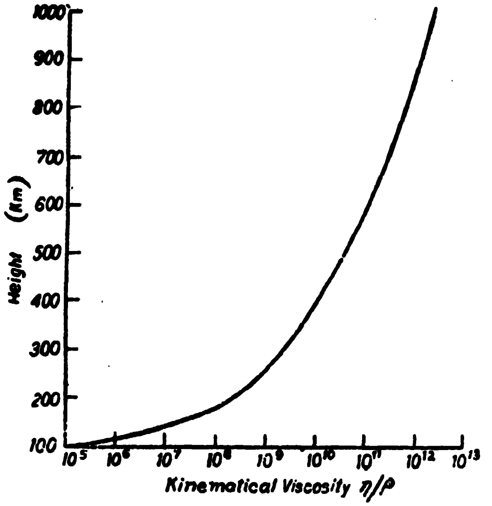


FIG. 3. The effective drag of one layer of gas upon another depends on the kinematical viscosity η/ρ (η—viscosity, ρ—density). The drag is thus greater in the higher than in the lower regions of the atmosphere.

opposite to the direction of rotation of the earth; the longitude at which the molecule will return to the atmosphere will be different from that at which it left.

(e) Effect of ionization

It has been mentioned that a neutral particle requires a velocity of about 11 km./sec. to overcome the pull of gravity and escape from the earth's atmosphere. We have not considered, however, the possibility of the atoms and molecules in the fringe region being ionized by the ultraviolet radiation of the sun. The motion and distribution of the particles, if ionized, will be profoundly influenced by the earth's magnetic field. The critical velocity for escape of an ionized particle and the variation of this velocity with latitude may be estimated after the method developed by Störmer [12] for calculating the trajectories of charged particles coming from infinity towards a magnetic dipole [see Chapter IX]. Störmer's calculations show that there is a space Q_γ, characterized by a certain value of γ, an integration constant, within which the charged particles cannot enter. For $\gamma < -1$, no charged corpuscle can reach the magnetic dipole. For γ lying between -1 and 0, charged particles from infinity may reach the magnetic dipole. There is, however, a toroidal space round the dipole where they cannot enter. The meridional curve of this space is given by

$$r = \frac{\sqrt{\gamma_1^2 + \sin^3\chi} - \gamma_1}{\sin\chi} \quad \ldots \quad \ldots \quad \ldots \quad (28)$$

where $\gamma_1 = -\gamma$,

χ—the angle which the radius vector makes with the magnetic axis Z,

r—the radius vector measured in units of length $\sqrt{M/HR}$, where M is the magnetic moment of the earth-magnet, H the intensity of the field at the point of observation and R the radius of curvature of the helical path of the charged particle so that $HR = mv/e$.

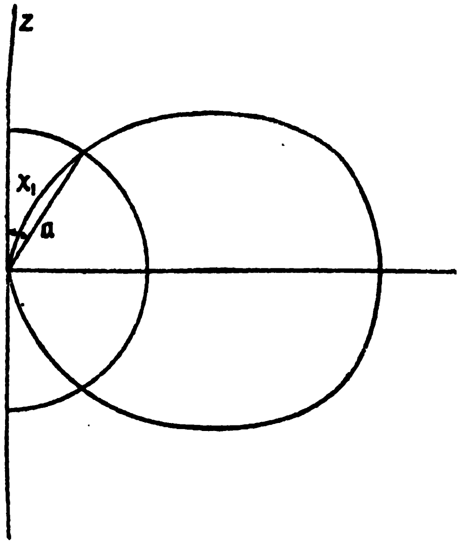


FIG. 4. Illustrating how the escape of charged particles from the upper atmosphere is restricted by the magnetic field of the earth.

The maximum angular distance χ_1 of the zone within which the charged particles may enter the earth's atmosphere (the earth is regarded as a small sphere of radius a placed round the magnetic dipole) is obtained by finding the intersection of the toroidal surface with a sphere of radius a. This is shown in Fig. 4. From Eq. (28) χ_1 is given by

$$\sin \chi_1 = \sqrt{2\gamma_1 a}$$

where a is measured in unit of length $\sqrt{M/HR}$.

Or, since the maximum value of γ_1 is 1,

$$\sin \chi_1 = \frac{(2a)^{\frac{1}{2}}(HR)^{\frac{1}{2}}}{(M)^{\frac{1}{2}}}, \text{ where } a \text{ is measured in cm.}$$

The case of charged particles escaping from the earth's atmosphere being exactly complementary to that of Störmer's, we can calculate the value of χ_1 giving the zone from which a charged particle having a given velocity v can escape from the earth's atmosphere to infinity. Table VI is prepared following the above method, and shows that the velocity required by an ion of atomic oxygen to disentangle itself from the earth's magnetic field is much greater than the velocity of escape of a neutral atom. Whereas the required velocity for a neutral particle is independent of the latitude, that for a charged particle is not so. The critical velocity increases with decreasing latitude. The radius a is taken as the distance from the centre of the earth to the height of the region of escape.

TABLE VI

Co-latitude (in degrees)	$HR = mv/e$	$\sqrt{M/HR}$ (cm.)	Velocity of escape (cm./sec.)
5·50	$3 \cdot 34 \times 10^3$	$1 \cdot 59 \times 10^{11}$	2×10^6
8·30	$1 \cdot 67 \times 10^4$	$7 \cdot 09 \times 10^{10}$	10^7
14·80	$1 \cdot 67 \times 10^5$	$2 \cdot 24 \times 10^{10}$	10^8
27·20	$1 \cdot 67 \times 10^6$	$7 \cdot 09 \times 10^9$	10^9
54·35	$1 \cdot 67 \times 10^7$	$2 \cdot 24 \times 10^9$	10^{10}
90	$4 \cdot 18 \times 10^7$	$1 \cdot 42 \times 10^9$	$2 \cdot 5 \times 10^{10}$

It should be noted that in the above calculation, the effect of gravitational pull on the charged particles has been neglected. If this is taken into account then under the combined action of gravitation and terrestrial magnetic field the particles would follow more complicated paths and the results obtained will be modified [13].

CHAPTER II

ATMOSPHERIC OSCILLATIONS—TIDES IN THE HIGH ATMOSPHERE

1. INTRODUCTION

Examination of the chart of daily pressure variation of the atmosphere shows that there are broad sinusoidal fluctuations in 24 hours with maxima, one at about 10 o'clock in the morning, and the other at the same hour in the evening (Fig. 1). The phase of the oscillation is approximately the same at all places at the same hour of the local time; its magnitude, from maximum to minimum, is about 2 mm. of mercury, that is, the total variation is about 2·6 parts in 1,000. Though small in magnitude its world-wide character is so persistent that even two centuries ago it did not escape detection by the physicists [1]. For a long time some

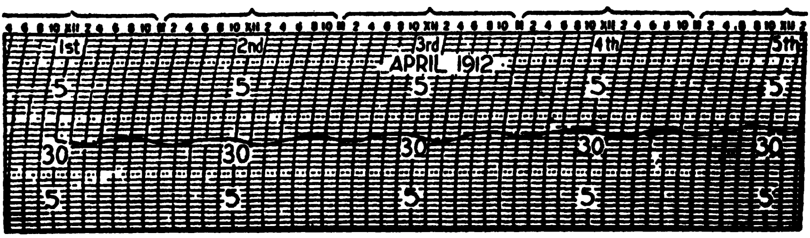


FIG. 1. Barogram depicting rhythmical diurnal variation of pressure. Two broad maxima may be noticed, one at 10 o'clock in the morning and the other at about the same hour in the evening. (The barogram is from Grand Turk Island, West Indies, April 1 to 5, 1912.)

aspects of the phenomenon had seemed mysterious, and, it is only in recent years that their satisfactory explanation has been obtained. Incidentally, the explanation has brought out the remarkable fact that the middle and the high atmospheres are seats of intense periodic winds having velocities approaching or exceeding 200 km. per hour. The wind system is world-wide in its character and is caused by the development of tidal forces due to the sun and is magnified manyfold by a resonance effect to be presently discussed. It extends up to the highest regions of the atmosphere and is responsible (besides pressure variations at the ground) for the so-called quiet day variation of terrestrial magnetic forces and analogous variations in the heights and ionization densities of the ionospheric regions. In the present chapter we are concerned mainly with the phenomenon of pressure variation. The ionospheric and terrestrial magnetic variations will be discussed in appropriate places in Chapters VI and VII respectively.

2. DESCRIPTION OF THE OSCILLATION PHENOMENA: HARMONIC ANALYSIS OF THE BAROGRAM

If the daily pressure chart, after certain necessary corrections, be subjected to harmonic analysis it is found that the pressure variation can be well represented by a few terms only of the series

$$\sum_n C_n \sin(nt+\alpha_n),$$

where t is the time measured from midnight to midnight, so that 1 hour of time is equivalent to 15°. The total pressure fluctuation may thus be regarded as made up of a few sinusoidal pressure vibrations, the amplitude of the nth vibration $C_n \sin(nt+\alpha_n)$ being C_n, its period $360/(n.15)$ hours and its phase α_n. Thus $n = 1$ represents a 24-hourly fluctuation, $n = 2$ a 12-hourly fluctuation, and so on. Actual harmonic analysis shows the existence of two main periods of 24 and 12 hours, together with two other small subsidiary periods of 8 and 6 hours. Fig. 2 depicts the relative phases, amplitudes and periods of the two main oscillations—the diurnal and the semi-diurnal [2].

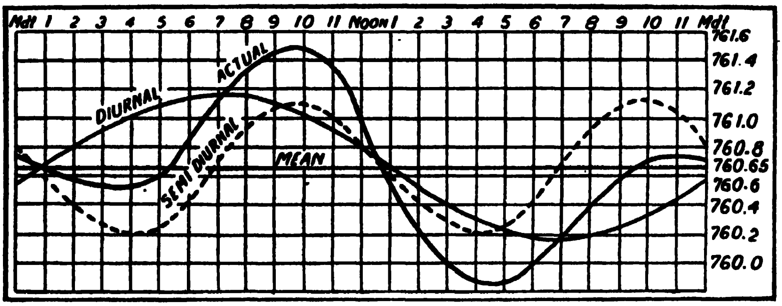


FIG. 2. Illustrating the diurnal and the semi-diurnal components of the average daily barometric variation. (The actual variation is for Washington D.C.) (After Bennett.)

(a) Diurnal wave

The diurnal wave is well marked and regular over equatorial oceans and tropical islands, but has not the same universal character as the semi-diurnal wave. It shows considerable variations in magnitude and phase in high latitudes, especially according as the country is plane or mountainous. For instance, near the equator where it is rather regular, it can be represented by $p_1 = 0{\cdot}30 \sin(t+0)$, whereas at Kalvesa on the Hungarian plane (47°N.) by $0{\cdot}36 \sin(t+340°)$ and at Bozen (about 600 ft. high, 47°N.) by $1{\cdot}38 \sin(t+19°)$, the unit in each case being 1 mm. of mercury. As the form shows, it is a progressive wave travelling from east to west round the earth with the sun.

(b) Semi-diurnal wave

The more important pressure oscillation is the semi-diurnal one. It shows considerable uniformity with regard to amplitude and phase. An

important characteristic of this oscillation is that it is resolvable into two components with different properties. According to Simpson [3], the entire semi-diurnal pressure wave comprising the two components, is expressible as

$$p_2 = 0{\cdot}937 \sin^3\theta \sin(2t+154^\circ)+0{\cdot}137\left(\cos^2\theta-\frac{1}{3}\right)\sin\{2(t-\lambda)+105^\circ\}$$
$$=p_2^{\,t}+p_2^{\,s},$$

where θ is the co-latitude of the place, and λ its longitude east of Greenwich; the unit, as before, is 1 mm. of mercury. The first component makes the maximum contribution to pressure oscillation and is a progressive wave (like the diurnal oscillation) travelling round the earth in 12 hours. It has a latitude effect represented by $\sin^3\theta$ law, so that the oscillation disappears entirely at the poles. The second component, at any instant of time, is a stationary wave of which the nodal circles are given by $\cos^2\theta = \frac{1}{3}$, i.e. 35° north and south of equator. Towards the pole from this circle there is maximum pressure at 11·5 hrs. G.M.T., since $\sin\{(2\times 11{\cdot}5\times 15)+105\} = +1$ and towards equator the pressure is minimum at the same time. It is to be noted that in the expression for the time-angle of the standing wave, the so-called world time $(t-\lambda)$ is constant over the whole world, t being the local time for any place. At any instant therefore, the pressure due to the standing wave is same at all points over a parallel of latitude.

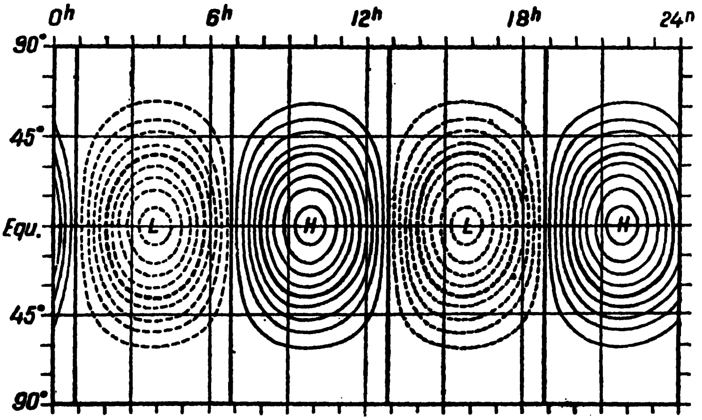


FIG. 3. Illustrating the progressive component of the semi-diurnal pressure wave in cylindrical projection. Longitudes have been indicated by local times. The lines indicate regions of same pressure—Isobars. Continuous lines are for high pressure and broken lines for low pressure. The distances between the lines correspond to difference of pressure of 0·1 mm. The waves travel from east to west. (After Bartels.)

The two components are illustrated in Figs. 3 and 4 and in cylindrical projection. (A circular cylinder is supposed to surround the earth touching along the equator, and the points on the earth are projected on this cylinder by lines parallel to the equatorial plane and ultimately the cylinder is cut along the generating line and laid on the plane of paper.)

The formula of Simpson for pressure variation with time and geographical location agrees closely with observed data. Other formulae

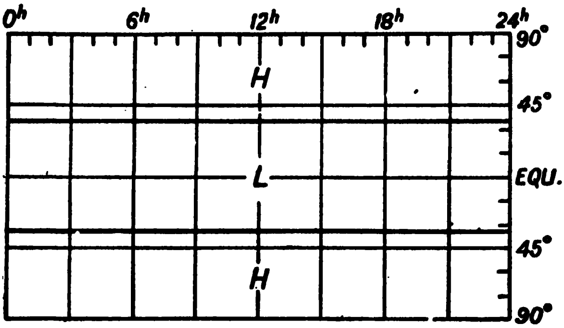


FIG. 4. Illustrating the standing wave component of the semi-diurnal pressure wave. The condition depicted is for the world time 11·5 hrs. After three hours the pressure variation will disappear. After 6 hrs. the regions of high and low pressures will be reversed. (After Bartels.)

have also been proposed either on theoretical grounds (for the travelling component) or for obtaining better agreement (for the standing wave component). Thus, for the standing wave component Wilkes [4] prefers the following empirical formula which fits the observed data better,

$$(0{\cdot}07-0{\cdot}1\,|\cos\theta|)\sin 2\,(t-\lambda)+0{\cdot}075\,|\cos\theta|\cos 2\,(t-\lambda).$$

For the amplitude of the travelling wave an expression involving spherical harmonics P_2^2 and P_4^2 of the order 2 and degrees 2 and 4 respectively has been suggested by Adolf Schmidt [5]. This is more sound from the theoretical point of view. The expression of Schmidt is as follows:—

$$\sin^2\theta(\cdot988-\cdot573\cos^2\theta)\sin(2t+154°)$$
$$=1{\cdot}046\,(P_2^2-0{\cdot}140\,P_4^2)\sin(2t+154°)$$
$$=[AP_2^2-BP_4^2]\sin\{(2t'+\lambda)+154°\}$$

where, $A=1{\cdot}046$, $B=0{\cdot}1464$, $t'=t-\lambda$,
t'—Greenwich time, t—local time, λ—longitude east of Greenwich.

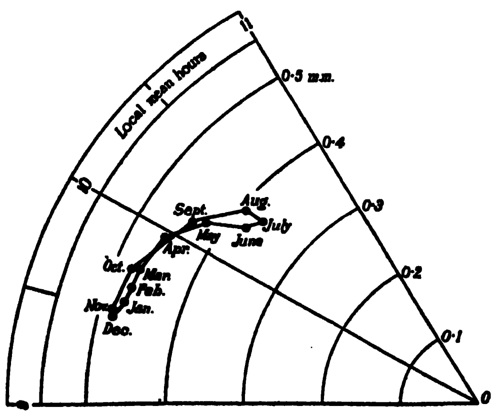
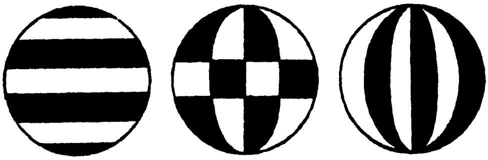


FIG. 5. Illustrating seasonal variation of the semi-diurnal pressure oscillation for Washington (D.C.). (After Chapman.)

It is to be mentioned that the semi-diurnal component possesses an annual variation which, though small, is fairly regular. This is illustrated in Fig. 5, prepared from the records of Washington, D.C. This method of depicting the phase and amplitude of the tidal oscillation vector is due to Bartels and is called by him *harmonic dial*.

NOTE. Fig. 6 illustrates how spherical harmonic functions P_n^l (cos θ) cos $m\phi$ may be used to represent atmospheric oscillations with nodes distributed over the surface of the earth. The representation is for a hemisphere and is in orthogonal projection. In (*a*) oscillations in θ corresponding to P_6^0 (cos θ) are represented. There are six



(*a*) (*b*) (*c*)

FIG. 6. Illustrating how spherical harmonic functions may be used to represent atmospheric oscillations with nodes distributed over the surface of the earth. The oscillations are in opposite phases in the black and the white portions.

nodal lines along which the oscillation amplitude is zero. The oscillations are in opposite phases in the black and the white portions. In (*c*) P_6^6 cos θ cos 6ϕ is represented. Since P_6^6 (cos θ) = 1, there are oscillations in ϕ only. Here again there are six nodal lines, and, as before, the ϕ-oscillations are in opposite phases in the white and the black segments. In (*b*) P_6^4 (cos θ) cos 4ϕ is represented. There are oscillations in both θ and ϕ. The actual oscillations may be represented by the superposition of such spherical harmonic functions with appropriate amplitudes and phases.

Minor waves: There is also an 8-hourly oscillation of about 0·1 mm. amplitude quite regular like the semi-diurnal oscillation. It has opposite phases in summer and winter as also in northern and southern hemispheres. Besides this an extremely small 6-hourly oscillation is also traceable.

(*c*) Air motions

The pressure distribution shown in Fig. 3 generates world-wide wind system as shown in Fig. 7. As will be seen from the figure the magnitude of the velocity is small compared to that of irregular winds in the troposphere. But, though small, the existence of these tidal winds has been established even for the lunar tide which is extremely small [4*a*]. However, as we shall see later there may be wind systems corresponding to these in the upper atmosphere of considerable velocity which are responsible for the well-known geomagnetic effect—the semi-diurnal variation of the intensity of the terrestrial magnetic field. It is also possible that the high velocities observed of the noctilucent clouds are due to this wind system.

The wind system shown in Fig. 7 is derived from the expression for the semi-diurnal pressure wave as follows: The equations for small atmospheric

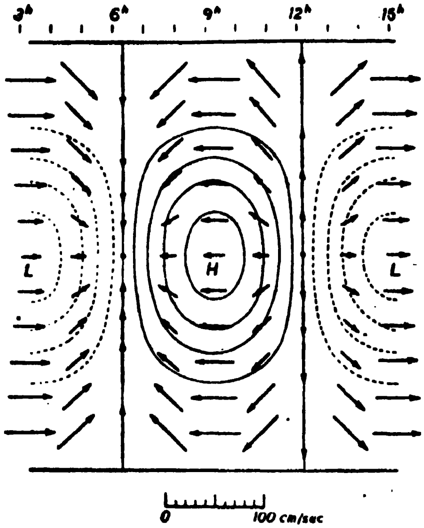


FIG. 7. World-wide wind system produced by the progressive pressure wave depicted in Fig. 3. The distances between the isobars correspond to 0·2 mm. pressure. The lengths of the arrows are proportional to the wind velocity according to the scale drawn under the figure. (After Bartels.)

motions in the horizontal direction (neglecting vertical motions) by forces due to the pressure gradient and the rotation of the earth are,

$$\rho_0\left(\frac{\partial u}{\partial t}-2\omega v\cos\theta\right)=-\frac{1}{a}\frac{\partial p}{\partial\theta} \quad .. \quad .. \quad (1)$$

$$\rho_0\left(\frac{\partial v}{\partial t}+2\omega u\cos\theta\right)=-\frac{1}{a\sin\theta}\frac{\partial p}{\partial\lambda} \quad .. \quad .. \quad (2)$$

where

ρ_0—average air density near ground,

u—velocity component towards south,

v—velocity component towards east,

ω—angular velocity of rotation of the earth $=7{\cdot}29\times10^{-5}$ radian/sec.,

a—radius of the earth,

θ—co-latitude,

λ—longitude.

$p=p_0+p'$, p_0 being the average pressure and p' the change due to the semi-diurnal pressure wave.

Now the expression for $p'=0{\cdot}937\sin^3\theta\sin\{2(\omega t+\lambda)+154^\circ\}$
$=A\sin\{2(\omega t+\lambda)+154^\circ\}$, say [page 33].

$$\therefore\ \frac{\partial p}{\partial\theta}=\frac{\partial A}{\partial\theta}\sin\{2(\omega t+\lambda)+154^\circ\}$$

$$\frac{\partial p}{\partial\lambda}=2A\cos\{2(\omega t+\lambda)+154^\circ\}.$$

Also

$$\rho_0 = \frac{p_0 m}{kT_0}.$$

Substituting the expressions for $\partial p/\partial\theta$, $\partial p/\partial\lambda$ and ρ_0 and putting f for $(kT_0/mp_0a\omega) \,.\, \partial A/\partial\theta$ and g for $2kT_0A/mp_0a\omega \sin\theta$ in Eqs. (1) and (2), we have, on integration

$$u = \frac{f+g\cos\theta}{2\sin^2\theta}\cos\left\{2(\omega t+\lambda)+154°\right\}$$

and
$$v = -\frac{g+f\cos\theta}{2\sin^2\theta}\sin\left\{2(\omega t+\lambda)+154°\right\}.$$

From these expressions of u and v, the wind system depicted in Fig. 7 can be determined. It should be remembered that the positive values of u and v mean north wind and west wind respectively. It may be noted in the figure that the velocity increases with the increase of latitude. This is an effect of the rotation of the earth. If this were not taken into account the velocity would have decreased with increasing latitude. For the equatorial region, however, the velocity is almost independent of the rotation of the earth.

3. THE PROBLEM OF OSCILLATIONS

(a) Introduction

The explanation for the pressure oscillations is to be sought in the vibration of the gaseous layer surrounding the globe forming its atmosphere. We picture an ocean of air surrounding the globe to a certain uniform depth which is in equilibrium due to the gravitational attraction of the earth. This system being stable will have, when disturbed, its own peculiar modes of vibration, the periods depending on the physical characteristics of the atmospheric layer. These are known as *free periods* of oscillation. In addition to this, the atmosphere will also have *forced oscillations* if it is subjected to periodic disturbances from external causes, such as the tidal action of the sun and the moon, and heating by the sun. According to the theory of forced vibration of a dynamical system, the atmosphere in such cases will be forced to vibrate in synchronism with the external periodic influence. Further, if any of such forced oscillations happens to possess a period identical with a free period of oscillation of the atmosphere, then, due to resonance effect, the oscillations will be strengthened to very high amplitude. This phenomenon of resonance, though comparatively rare for natural systems, appears to play an important rôle in atmospheric oscillations.

(b) Diurnal oscillations; periodic heating by the sun

We can form an idea of how the diurnal oscillation is produced by the heating effect of the sun. The heating and the consequent expansion and vertical rise of the air on the sunlit side of the earth on the one hand, and the simultaneous cooling and fall on the dark side on the other, establish

a pressure gradient from the warmer to the cooler side at all levels. This pressure gradient tends to produce maximum pressure in the coldest and a minimum pressure in the warmest regions. Now since the cold and the hot regions are perpetually moving round the earth once every 24 hours, the pressure gradient described above exists perpetually trying to establish an equilibrium and causes the diurnal pressure variation.

(c) Solar semi-diurnal oscillations

Solar semi-diurnal oscillation is caused partly by the tidal action and partly by the heating effect of the sun. According to Chapman [5a], the contribution of the thermal action is at least of the same order of magnitude as that of the tidal action. The mode of operation of the former is comparatively simple; that of the latter is more complicated and had for a long time been a puzzle to the mathematical physicists.

(i) *Effect of solar heating.*—The temperature variation of the atmosphere has, besides the diurnal, a semi-diurnal component. According to Chapman, the expression for the semi-diurnal variation of temperature, as based on observational data, may be taken as

$$0°{\cdot}4 \sin (2t+80°).$$

This semi-diurnal temperature variation produces semi-diurnal pressure oscillation which leads it by about 135°. The temperature variation is partly the cause and partly the effect of the semi-diurnal pressure oscillation. There are certain direct causes which tend to produce a temperature variation in the atmosphere. This temperature variation gives rise to atmospheric motion. 'The changes of volume associated with these motions themselves produce changes in temperature, and the observed variation of temperature is a kind of resultant of these two less and more direct effects of the solar radiation falling upon the earth.'

As mentioned above, the phase and amplitude of the observed semi-diurnal pressure oscillation are determined by the combination of the oscillations due to the semi-diurnal temperature variation and the gravitational tidal oscillations of same period to be discussed in the following sections.

(ii) *Tide-generating forces—tidal oscillations of oceans.*—The tide-generating forces which produce a part of the semi-diurnal pressure oscillation are exactly the same as those producing tides in oceans. It will be helpful to first recall briefly how these forces are developed.

Let us consider the lunar tide. The moon and the earth's centre revolve round their common centre of gravity in one month. The centre of gravity is situated on the line joining the two centres at the average distances of 3,000 miles from the earth's centre and 237,000 miles from the moon's centre. The earth is about 80 times as heavy as the moon and they are held at their respective distances by the balance of the centrifugal and the centripetal forces. The moon in her revolution always presents the same face towards the earth and, as such, completes one rotation round her axis in course of one revolution. The earth, however (if, for the moment

we ignore its diurnal rotation) revolves round the common centre of gravity without any corresponding rotation. Hence, every part of the earth describes its own circle of revolution and is thus subject to equal and parallel centrifugal forces. (One can imagine, for instance, a wheel with spokes revolving in its own plane with the spokes always remaining parallel to themselves.) Now, the total centrifugal force acting on the whole earth is just such as to balance the total of the centripetal forces due to the moon's attraction. But though there is balance on the whole, there is not balance at every part. This is because, though the centrifugal forces acting on the earth are uniformly distributed, being equal and parallel at every part of the earth, the moon's attraction on the earth is not so, being greater for the parts nearer than for those remoter. Hence, at the nearer parts of the earth the moon's attraction overbalances the centrifugal forces and at the remoter parts the centrifugal forces overbalance the attraction. These overbalances are the tide-generating forces.

The magnitude of the tide-generating force at any point at the earth varies inversely as the cube of the distance from the moon. For example, while the average pull of the moon at any point of the earth (on unit mass), is equal to that at its centre, namely GM/D^2, that at A, the nearest point, is $GM/(D-a)^2$, where G is the constant of gravitation, M the mass of the moon, D the earth-moon distance and a is the radius of the earth. The difference of the accelerations is equal to

$$\frac{GM}{(D-a)^2}-\frac{GM}{D^2}\doteq\frac{2aGM}{D^3}$$

since, $D\gg a$. Hence the moon's attraction at A overbalances the centrifugal force at this point (which is equal to the average pull of the moon) by this amount. Similarly at B, the point farthest away, the centrifugal force overbalances the moon's attraction by the same amount.

In Fig. 8 the tide-generating forces calculated in the same way, for a number of points on the surface of the earth, for a section passing through the moon, are depicted. It will be noticed that the force is wholly outwards at A and B and wholly inwards at the ends of the diameter at right angles to AB. At other points it has a component parallel to the surface. The outward and inward forces have merely the effects of slightly decreasing or increasing the attractions of gravity at these points. At the other points the forces produce horizontal flow of water towards the two antipodal points A and B tending to make the surface of the ocean assume the flattened form shown in Fig. 10 (assuming the water to be perfectly mobile and free to take up that form quietly). In Fig. 9 is depicted how the horizontal tide-generating force is distributed on the surface of the earth. At A the moon is vertically overhead. At places on the dashed circle the moon is on the horizon. P is the north pole. The line O_1O_2 represents a latitude. Since O_1O_2 cuts the arrows unsymmetrically, an observer carried along O_1O_2 in course of the diurnal rotation of the earth observes an unsymmetrical rise and fall of water.

Mathematically the tidal action on a rotating globe is much more complicated than on a fixed globe.

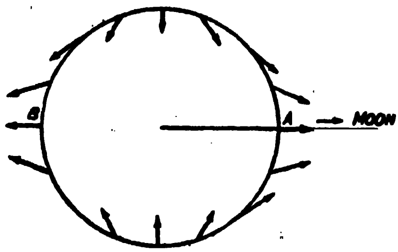


FIG. 8. Illustrating the tide-generating forces. The circle is a section of the surface of the earth passing through the moon. The forces are outwards at A and B and inwards along the diagonal at right angles to AB. At other points the forces have horizontal components. These cause horizontal flow of water towards the two antipodal points A and B.

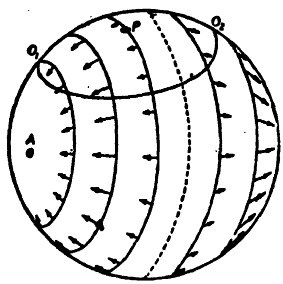


FIG. 9. Illustrating the distribution of the tide-generating forces over the surface of the earth. The moon is vertically overhead at A. At places on the dashed circle the moon is on the horizon. P is the north pole. The line O_1O_2 represents a latitude. An observer carried along O_1O_2 will experience an unsymmetrical rise and fall of water. (See Fig. 10.)

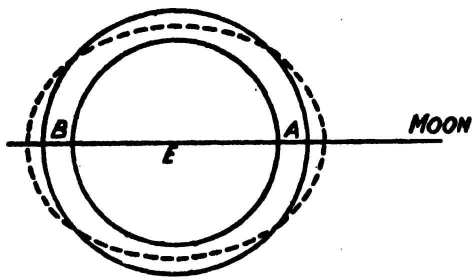


FIG. 10. Illustrating how as a result of the flow of water caused by the tide-generating forces depicted in Figs. 8 and 9, the spherical mass of water (outer circle in continuous line) assumes the flattened form shown by the broken line.

If the potential of the tidal force of the sun or the moon be calculated, it is found to contain three terms. The first term is very slowly variable with fortnightly (for moon) and six-monthly (for sun) periods; the two others can be represented by

$$A \cos t + B \cos 2t,$$

in which A and B involve the latitude of the place and declination of the disturbing body (for this latter A and B are also very slowly varying), and t is the hour angle of the disturbing body. The first of these has a period of 24 hours for the sun and 1 lunar day for the moon while the second has a period of 12 hours for the sun and 1/2 lunar day for the moon. These two terms are represented in the tidal force and are responsible for the diurnal and semi-diurnal tides, as they set up forced vibrations of the ocean.

Tidal waves, also called *long waves*, have the property that the wavelengths are large compared to the depth of water. The velocity of long waves is independent of the wavelength and is equal to $\sqrt{gH}$, where H is the depth of water.

(iii) *Atmospheric tides.*—The above considerations of tidal disturbance are also applicable to the forced oscillations of the atmosphere caused by the action of the sun and moon. The component of the barometric oscillation represented by p_1 [Sec. 2(*a*)] evidently involves a diurnal tide. The first term $p_2{}^t$ in the expression for the semi-diurnal oscillations [Sec. 2(*b*)] representing a 12-hourly wave travelling from east to west involves the semi-diurnal tide. The origin of the second term $p_2{}^s$—representing a stationary wave is still obscure. Now, since the sun and the moon both exert tidal influence, the semi-diurnal wave should have components of both solar and lunar periods, i.e., there should be semi-(lunar)diurnal and semi-(solar)-diurnal waves. But while the solar semi-diurnal component is easily recognizable, the lunar semi-diurnal component is so weak that Laplace looked for it in vain. It has an amplitude which is only one-sixteenth of the solar tide [see Sec. 5]. This raises great difficulty. Since the tide-generating force of the disturbing body, as stated before, varies as M/D^3, the moon on account of its proximity to earth exerts, in spite of its smaller mass, a tide-generating force which is nearly 2·5 times as powerful as that of the sun. The lunar component of the atmospheric tide ought, therefore, to be 2·5 times stronger than the solar component. Actually however, the reverse is the case; the observed amplitude of the solar semi-diurnal oscillation is about 100 times greater than what it ought to be. This difficulty had long stood in the way of the acceptance of the tidal-*cum*-resonance theory, first put forward by Kelvin [6], in which it is assumed that the atmosphere has a natural period of free oscillation of 12 hours due to which the oscillations produced by the feeble tidal forces of the sun are magnified hundredfold. The resonance theory will now be discussed and it will be shown how recent investigations have indicated the way out of the difficulty.

4. ATMOSPHERIC TIDAL OSCILLATIONS—TAYLOR-PEKERIS THEORY

As mentioned above, the basic assumption of the resonance theory is that the atmosphere has a natural period of oscillation of 12 hours. Theory, in fact, demands that for a nearly hundredfold magnification of the tidal

oscillation the free period should be within 6 minutes of 12 hours. It is therefore first necessary to enquire if the atmosphere has a mode of free oscillation of this period. Now, on the assumption of an isothermal or adiabatic temperature distribution, the period of free oscillation of the atmosphere can be computed from the speed of explosion in air. Such computations have been made from the propagation of pressure waves sent out by the Karakatau eruption in 1883 [8, 9] when the waves cut round the earth several times (Fig. 11) and also from the waves produced

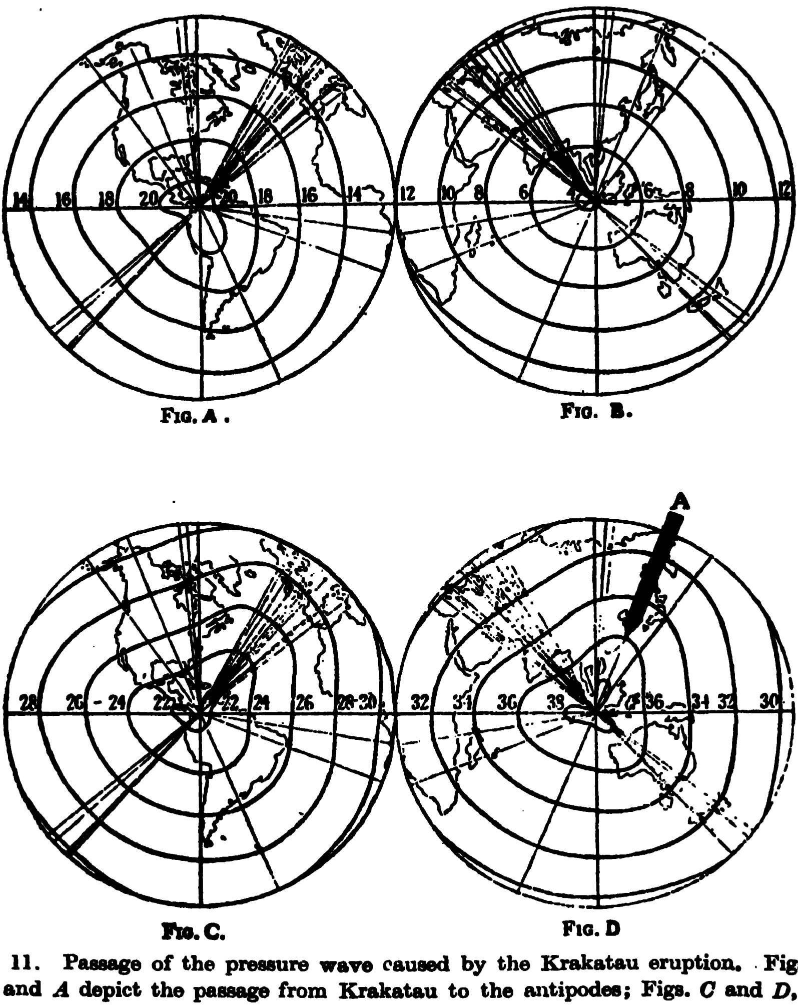


FIG. A. FIG. B.

FIG. C. FIG. D

FIG. 11. Passage of the pressure wave caused by the Krakatau eruption. Figs. *B* and *A* depict the passage from Krakatau to the antipodes; Figs. *C* and *D*, the passage from the antipodes back to Krakatau. The figures along the abscissa indicate the hours after the explosion. The arrow *A* indicates the strong asymmetry in the curves of equal pressure which has developed after it has travelled to the antipodes and back. This is presumably due to the effect of world-wide air currents. (After Strachey.)

by the fall of the Great Siberian Meteor of June 30, 1908 [10]. For the former the speed of the so-called long waves produced in the ocean of air

was found to be 318·8 m./sec. For the latter, not determined with great accuracy, it was found to be 318 m./sec. These speeds correspond to a period of about 10·5 hours for the free oscillations. It is clear that as this period does not lie within the narrow margin of 6 minutes of 12 hours, the corresponding mode of oscillations cannot contribute to the tidal resonance phenomena as envisaged by Kelvin. (It is interesting to note in this connection that the total energy of the atmospheric wave excited by the Krakatau eruption is estimated by Pekeris to be 10^{24} ergs. Compared to this the energy of explosion of the atom bomb over Nagasaki in 1945 was equivalent to that of 20,000 tons of T.N.T. which is 10^{21} ergs. It is not known what fraction of this energy excited long wave of the type produced by the Krakatau eruption.)

New light was thrown on the problem of solar semi-diurnal oscillations by the work of Taylor [11] followed by that of Pekeris [12] when it was shown that it is only in the case of isothermal and adiabatic equilibrium (as assumed by previous workers) that the atmosphere has only one mode of oscillation with one free period. For an arbitrary distribution of temperature there are, in general, more than one mode of oscillation and more than one free period. By assuming a temperature distribution of the type *ABCDEFG* in Fig. 12 Pekeris showed that there is a mode of oscillation

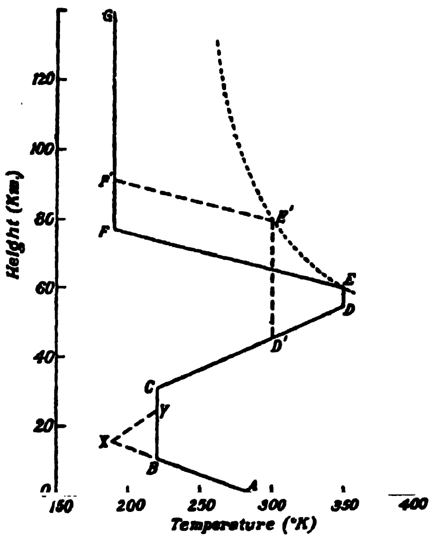


FIG. 12. Idealized temperature distribution with height (based on observations). For this type of temperature distribution the atmosphere has a mode of oscillation with a period of 12 solar hours. (After Wilkes.)

with a node at about 30 km. height and that the period of oscillations for this mode is 12 hours, namely, that required by the resonance theory. (In the distribution actually assumed by Pekeris the points D' and E coincided.) The grounds for assuming this distribution will be discussed presently. We will now give a brief account of the Taylor-Pekeris theory and its implications after Weekes and Wilkes [13]. (See also, Wilkes, *Oscillations of the Earth's Atmosphere*, Cambridge University Press, 1949.)

The energy which is introduced periodically into the atmosphere by the tide-raising forces (and also, in the case of the sun, by the thermal force) is generated mainly near the surface of the earth where the density of the air is greatest. This energy moves outwards as a sort of spherical wave motion. The motion is, however, more complicated than that of the sound wave, there being both horizontal and vertical motions of air. It can be shown that these waves may be regarded as plane waves propagated in a medium with a refractive index given by

$$\mu^2 = -\frac{1}{4}+\frac{1}{h}\left(\frac{dH}{dx}+\frac{\gamma-1}{\gamma}H\right),$$

where $x(= \int dz/H)$ is a function of the height z, γ is the ratio of specific heats and h is a constant and may be regarded as an equivalent depth of the atmosphere. The value of h depends both upon the mode and on the period of oscillation. For a particular mode the period decreases as h increases. Simple inspection of the expression for μ^2 shows that μ^2 can be negative (remembering that $H = kT/mg$) if the temperature is sufficiently low, or if there is a negative temperature gradient, or, with a combination of these two effects.

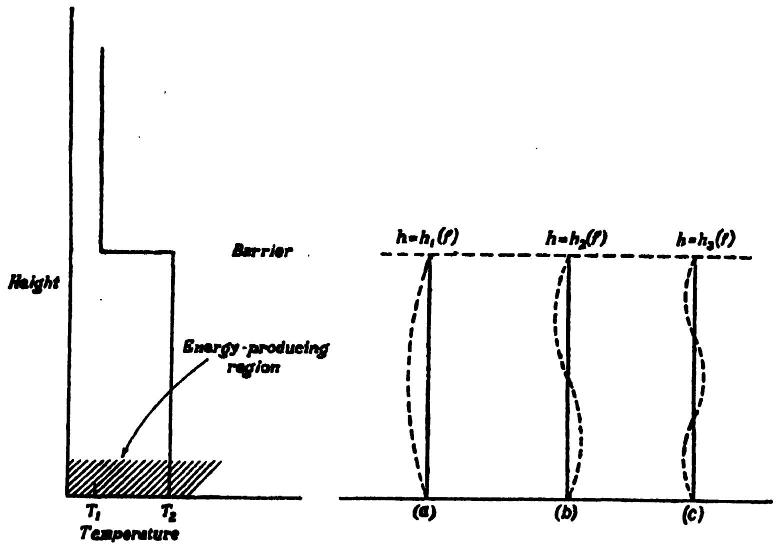


FIG. 13. Illustrating how energy introduced by tide-generating forces may be trapped by the presence of a region of low temperature in the upper atmosphere. (After Wilkes.)

Now consider a hypothetical atmosphere with temperature distribution as shown on the left in Fig. 13. There is a region of uniform temperature T_2 on which rests a region of lower temperature T_1. If T_1 is sufficiently low then, for a given value of h, μ^2 may be negative. Waves, as discussed above, of periods corresponding to this and all lower values of h will then be reflected back at the discontinuity. The tidal energy producing these waves

will therefore not be able to pass beyond this level which will act as a barrier. The energy will be 'trapped' as it were between this level and the ground where the energy is introduced. It is thus clear that if the height of the region T_2 is suitably adjusted in relation to the period then free oscillations as shown in (*a*) will occur. In analogy with oscillations in an organ pipe, oscillations of types (*b*) and (*c*) with one and two nodes respectively may also be expected to occur. The periods of these different modes are, however, not simply related to one another as in an organ pipe. Further, unlike oscillations in an organ pipe, the trapped oscillations cannot occur for all higher orders with increasing number of nodes. This is because as the number of nodes increases, the value of h decreases and for these lower values of h the barrier becomes transparent. In Fig. 13 for example, oscillations with one, two or three nodes may occur, but, ultimately, as the value of h becomes sufficiently decreased, the barrier ceases to be a barrier and the series of possible values of h comes to an end.

In the above example we have assumed that the region of low temperature T_2 (i.e., the region of refractive index zero or imaginary) extends up to infinity. For such case there is total reflection of the waves. If, however, the region of low temperature is of finite thickness, so that above and below it the refractive index is positive, then the barrier will be only partially reflecting and energy will leak through it. The transparency of the barrier depends upon its thickness. For large thickness, the barrier is only slightly transparent and resonance phenomena will occur; for small thickness the barrier may be almost wholly transparent and resonance may be so damped as to be unobservable.

In considering resonance we have neglected losses by degradation of the oscillation energy into random energy through the effects of viscosity and thermal conduction. This is justifiable for the lower and the middle atmosphere. But in the higher regions, in the ionosphere, for example, these losses become important. Atmospheric oscillation energy as may ultimately find its way to the very high regions of the atmosphere is absorbed there. It is, however, to be noted that insofar as damping of oscillations in the trapped region is concerned, this may be determined simply from a consideration of the thickness, i.e., leakiness of the barrier.

Let us now consider the temperature distribution curve *ABCDEFG* in Fig. 12. The distribution is an idealized one—based on actual observations. The drop in temperature in the troposphere (0–10 km.) and the constant temperature above (10–30 km.) are well known from sounding balloon observations. The rise in temperature *CD* in the region (30–50 km.) is inferred from abnormal sound propagation experiments (see Chapter IIIA). The gradient is taken as 5·5°K per km. The rise is followed by a fall *EF* up to 78 km. with the temperature gradient taken as 9°K per km. The temperature is then assumed to remain constant. A cold region above 78 km. is inferred from the presence of noctilucent clouds in the region and also from meteor data. Direct evidence of a drop in temperature in this region is now furnished by V-2 rocket experiments.

Calculations show that for this type of temperature distribution, for a value of h slightly greater than 10 km., there exists a mode of oscillation with period of 10·5 hours. This evidently is the mode excited by the Krakatau eruption and by the fall of the Great Siberian Meteor. h, calculated from the speed of the long wave (318·8 m./sec.) produced by the eruption, was found to have the value 10·4 km., that is, somewhat greater than 10 km. as deduced by Pekeris. For this mode of oscillation the cold top barrier consists of not only of the low temperature region in the stratosphere (CB) but also the region of falling temperature in the troposphere (AB). In fact, the trapping of energy begins right from the surface of the earth where the energy is introduced. Calculation further shows that for this barrier no other higher mode of free oscillation is possible, because the barrier becomes transparent for such modes.

It is, however, found that for the temperature distribution as in Fig. 12, the cold top EFG above the hot layer CDE acts as a second barrier, and that for this case, a mode of oscillation with h less than 10·4 km. is possible. In fact, there is a mode with a horizontal nodal surface at 30 km. for which $h = 7{\cdot}9$ km. The time period of this mode is 12 hours. This, obviously, is the much sought for oscillation of 12 hour period which is magnified by the semi-diurnal solar tidal force.

It is to be noted that for this second barrier, oscillations with no node are not possible, because for such oscillations, the lower stratosphere acts as barrier and energy introduced near the ground cannot leak through to the upper regions.

The velocity of the air particles from ground level up to 30 km. is small and nearly constant. At the nodal surface at 30 km. the velocity and pressure variations vanish. Above and below this level the oscillations are in opposite phases. The velocity above now rapidly increases with height and at 100 km. level (where it is nearly horizontal) it is about 200 times the velocity below 30 km. These results are depicted in Fig. 14 after Pekeris.

It should be mentioned that the action of the second barrier is due partly to the region of falling temperature EF and partly to the isothermal region FG. If it is assumed that the temperature does not rise again, the barrier will act as a perfect barrier. However, we know that the temperature above the isothermal region FG rises again due to absorption by oxygen gas (see Chapter XI). The barrier consisting partly of EF and partly of FG has thus a finite width and its effectiveness depends on the total effective width. Keeping EF the same, if FG is assumed to be 28 km. thick, then it is found that for $h = 7{\cdot}9$ km. amplification of nearly 150 is obtained over the equilibrium tide.

It is particularly to be noted that the assumed temperature distribution $ABCDEFG$ (Fig. 12), for which oscillations of a 12 hour period (with a nodal surface at 30 km.) exist is not at all critical [14]. It may be varied over a considerable range without much change in the value of the period. For example, it is found that keeping the gradients of the rising (CD) and falling (EF) temperatures the same, if the distribution with height is

altered by altering the height of ED, keeping E always on the dotted curve, the same values of h, namely 7·9 km. and 10·4 km., are obtained. The

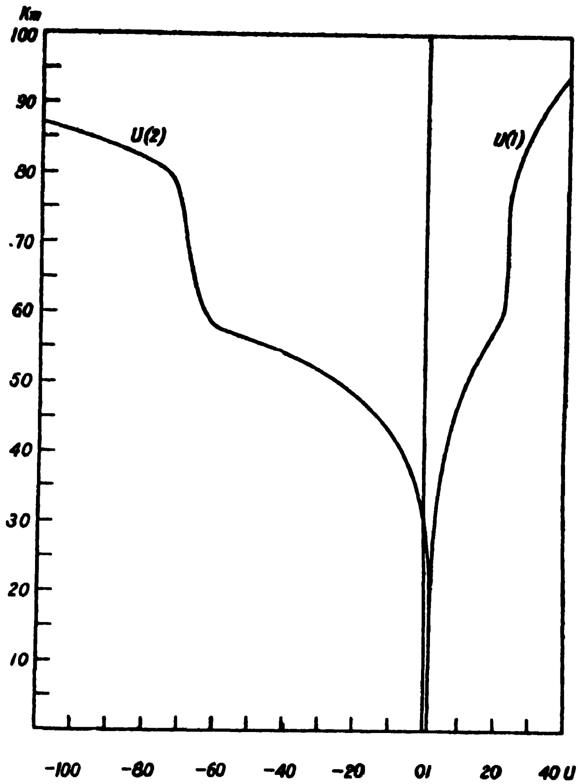


FIG. 14. Illustrating how the tidal velocities increase with height with respect to those at ground. Curve $U(1)$ for 10·5-hourly mode and curve $U(2)$ for the twelve-hourly mode. (After Pekeris.)

position of the dotted curve, of course, depends on the assumed temperature gradient. If, for instance, the rising portion CD is taken to start from 35 km., instead from 30 km., the dotted curve is shifted down parallel to that shown in the figure by 4 km. It is further found that the mode with $h = 7{\cdot}9$ km. (i.e., the 12 hour period of oscillation) is very little altered even if the temperature of the stratosphere (BC) is altered by $\pm 10°$C.

5. LUNAR SEMI-DIURNAL TIDAL OSCILLATIONS

The lunar semi-diurnal tidal oscillations, as already mentioned, are of very small amplitude, being only about $\frac{1}{16}$ of that of solar origin. Its determination is a difficult task particularly at temperate latitudes where the amplitude is extremely small, about a hundredth of a millimetre. In fact, many investigators among whom were famous men like Laplace and Airy, failed to detect it. The first successful determination of lunar tidal effect was made by Lefroy [15] at St. Helena in 1842, and later by Sabine

in 1847 [15]. Lefroy's success was principally due to his working with tropical data where the variations are comparatively larger. Lunar tidal oscillation in higher latitudes was first determined successfully by Chapman in 1918 with Greenwich data [16]. Chapman included in his analysis data of only the days on which the barometric fluctuation was less than 0·1 inch, and thus reduced greatly the random variations. His result showed that the amplitude at Greenwich was about 0·01 mm. This is a remarkable achievement when it is remembered that the amplitude of the tidal oscillation is less than the error of reading the barographic record! That a systematic variation of this magnitude can be determined provides an interesting illustration of the theory of random errors. The lunar tide is now known for 54 stations (including three in the tropics) mainly through the efforts of Chapman and his colleagues [16, 17, 18, 19].

The lunar tidal oscillations possess many interesting features. The amplitude is a maximum near the equator and generally decreases at higher latitudes. If the phases of the oscillations are averaged over several whole years, it is found that the maximum barometric pressure is attained nearly, though not quite, at the time of local lunar transit, i.e., when the moon crosses the meridian upper or lower. The difference of the time may amount to more than two hours. A remarkable and as yet unexplained characteristic of this phenomenon is that it undergoes a very marked annual

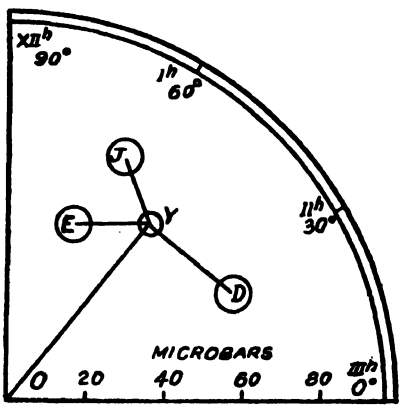


FIG. 15. Illustrating the variation throughout the year of the amplitude and phase of the tidal vector at Taihoku (Formosa, 25°N.). (After Chapman.)

variation [16, 20]. A typical example is shown in Fig. 15 for Taihoku (Formosa, 25°N.) for the period 1897–1932. The figures along the arc indicate the lunar hour, 12 hour corresponding to the time of local lunar transit. The length OY represents the average pressure while its direction gives the average hour of high tide. The circle round Y represents the probable error of the position of the end-point of the tidal vector OY. J, E, and D similarly indicate the positions and probable errors of the tidal vector in three different seasons of the year, J for the four months May–August round the June solstice, D for the four months November–February round the December solstice, E for the remaining months (equinoctial). The figure clearly shows the seasonal variation regarding both the strength and

the phase of the 'tidal vector'. As mentioned before the origin of this remarkable change is still unknown. It may further be mentioned that the change is not, what may be called, a seasonal one, i.e., it is not opposite in the two hemispheres. Since the tide generating force of the moon does not undergo any important annual change it follows that any large annual variation of the tide must be caused by an annual variation in the atmosphere which in its turn would depend on the season. That this is not so seems to imply that the atmospheric changes which cause the lunar atmospheric tide are not perceptible on the surface of the earth.

The lunar semi-diurnal period is off the resonance period of 12 hours by a significant amount. Hence it is easy to understand that the amplitude of the lunar tidal oscillation would be much smaller than the solar oscillation. If it is assumed that both the lunar and the solar tidal forces excite the same Taylor-Pekeris mode of oscillation then the variations of the amplitude and the phase of the two tidal oscillations with height may be compared.

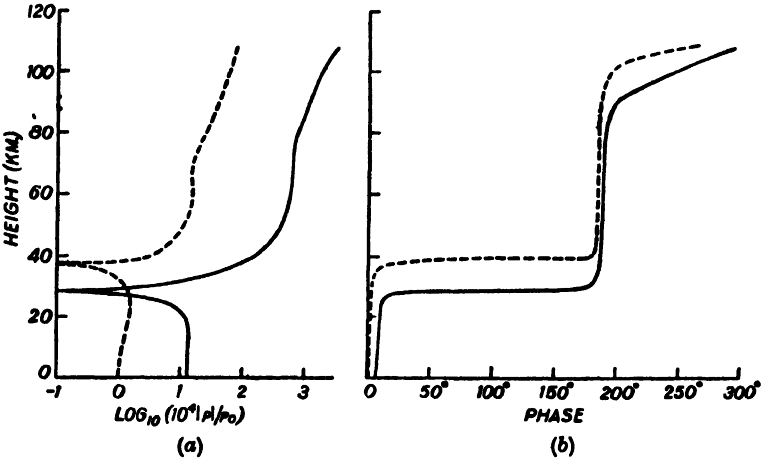


FIG. 16 (*a*), (*b*). Illustrating how the amplitudes of the lunar and solar semi-diurnal pressure oscillations vary with height assuming the same mode of oscillation for both. Dashed-curve—lunar oscillation, solid-line curve—solar oscillation. (After Wilkes.)

In Figs. 16(*a*) and (*b*) these variations are depicted, the dotted curves referring to lunar oscillation and the solid curves to the solar. It will be noticed that the two curves run nearly parallel to one another. The positions of the nodes in the amplitude curves are, as expected, different on account of the difference in the periods. It will also be noticed that the ratio of the amplitudes of the solar and the lunar variations in the upper regions is smaller than at the ground. Fig. 16(*b*) shows how the oscillations in the higher regions differ in phase by 180° from that at the ground. Both these results are verified from the comparison of the pressure oscillation at the ground with the solar and lunar quiet day magnetic variations which have their

origin in the solar and lunar tidal oscillations in the ionized region of the upper atmosphere. (See Chapter VII.)

Mention may be made of two other lunar tidal effects.

As the tide-raising force varies inversely as the cube of the distance and, as the distance of the moon from the earth is comparatively small, variation in lunar tide with the change of the lunar distance may be expected. (The earth-moon distance varies between 221,000 and 252,000 miles.) Lefroy looked for this effect in 1847 and found that though some stations showed the effect of the right order others did not. The matter requires further investigation.

The compressions and rarefactions of the atmosphere associated with the tidal oscillations are adiabatic as in the case of sound waves. This, though it may seem surprising because of the long period of tidal oscillations, is due to the fact that the tidal oscillations are on such a vast worldwide scale that heat transfer from one part to another is negligible. A small temperature variation accompanying the tidal oscillations is therefore to be expected particularly for the lunar tide, because, unlike for the solar tide, the temperature variation is not masked by any heating effect. If P_0 and T_0 are the static pressure and temperature, and P and T the semi-diurnal pressure and temperature due to tidal oscillation, then it is easily shown that

$$\frac{T}{P}=\frac{\gamma-1}{\gamma}\frac{T_0}{P_0},$$

where γ is the ratio of the specific heats. The existence of the temperature variation has been demonstrated by Chapman by analysis of 2-hourly data at Batavia for the period 1866 to 1928 [17]. His results are depicted in

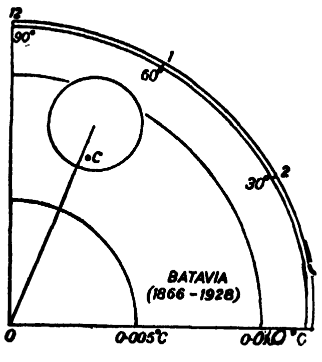


FIG. 17. Illustrating semi-diurnal temperature variations of the air temperature caused by lunar tidal oscillations. Point C is the theoretically computed point. (After Chapman.)

the harmonic dial Fig. 17. It is seen that the predicted position of the end-point of the temperature tidal vector lies within the limits of statistical expectation.

CHAPTER III

TEMPERATURE AND DENSITY DISTRIBUTION IN THE MIDDLE ATMOSPHERE

There are two useful indirect methods of determining the temperature and density distribution in the middle atmosphere. These are, studies of the abnormal propagation of sound waves and the meteoric phenomena. The present chapter is therefore divided into two sub-chapters one on each of the subjects. It is to be mentioned that meteoric phenomena also include the ionization produced along the meteor trails. Study of this aspect of meteoric phenomena will be taken up in Sec. 5 of the second part of this Chapter and also partly in Chapter VI on ionosphere.

IIIA—ANOMALOUS PROPAGATION OF SOUND

1. INTRODUCTION

It has been known for a long time that beyond the usual zone of audibility—60 km. to 80 km.—surrounding an explosion, there exists a second zone of audibility separated from the first by a zone of silence [1]. During the two Great Wars it was noticed by many people that gun-fire could be heard at distances ranging from 150–400 km., though persons nearer to the battlefront could not hear the sound [2].

Fig. 1 depicts the silent and the audible zones for an explosion which occurred at Oppau in Germany on September 21, 1921. It will be noticed that while sound was heard directly up to a distance of 100 km. surrounding the centre of explosion, the region between 100 and 200 km. was practically a zone of no audibility. There were a few points where sound was heard in this zone; but the majority of observers did not notice anything. Beyond 200 km. the sound was heard by practically all observers in the east and south.

The obvious explanation of this curious phenomenon is that the sound rays proceeding upwards in an oblique direction skip the silent zone and come back to the earth, being bent down by some process in the upper atmosphere. Regarding the nature of this process various suggestions have been put forward from time to time. Of these, the one that is universally accepted at present is that the bending is due to increased velocity of sound waves caused by a gradient of rising temperature in the region above 40 km. A critical study of this anomalous propagation phenomenon provides a method of estimating the temperature of the region of middle atmosphere (35 to 60 km.). The temperature-rise is due to absorption of solar radiation by atmospheric ozone [see Chapter IV]. Before discussing this hypothesis it would be useful to describe briefly two other possible explanations of the anomalous propagation phenomenon and to show how they fail.

Wind hypothesis.—It has been suggested that the bending may be due to a velocity gradient of wind at high altitudes. Such a gradient (velocity

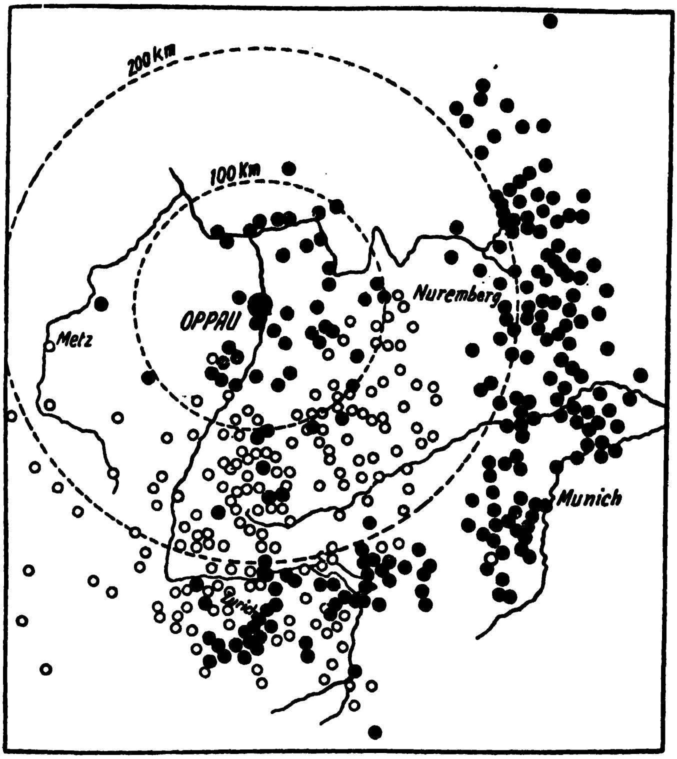


FIG. 1. Zones of audibility and of silence for the explosion at Oppau, Germany, on September 21, 1921. The black dots represent places where the sound was heard. The circles represent the positions of observers who did not hear the explosion Towards east and south, there is a zone of audibility of roughly 100 km. extent. The zone 100 to 200 km. is a zone of silence. Beyond 200 km. there is again a zone of audibility. (After Quervain.)

increasing upwards) will cause the direction of propagation of the waves—going upwards obliquely in the windward side—to bend down gradually. The wind hypothesis, however, does not explain all the observed features of the anomalous propagation phenomena. The audible and the silent zones, for instance, are roughly circular in shape indicating that the anomalous sound propagation occurs in all directions round the source. But according to the wind hypothesis the abnormal propagation should take place only in the direction in which the velocity has a component. Again, according to this hypothesis the abnormal propagation would be observed on windy days only; actually, however, it is observed equally on windy and on quiet days. Finally, though there are comparatively frequent changes of wind velocity at high altitudes, no corresponding changes of the inner radius of the abnormal audibility zones have been observed.

Possible existence of lighter gases in the middle atmosphere.—The velocity of sound being inversely proportional to $\sqrt{m}$ (see Eq. (1)), it is evident that a preponderance of lighter gases like hydrogen and/or helium from above, say, 30 to 40 km. would cause an increase in sound velocity in the middle atmosphere, leading to the anomalous propagation phenomena. The velocity of sound at such heights should then be higher than at the ground in the ratio of $\sqrt{m_{(\text{air})}/m_{(\text{light gas})}}$. Recent investigations show, however, that the atmospheric constituents up to about 80 km. are the same as near the ground and exist practically in the same proportion. (Ozone is present in the middle atmosphere but its amount is such and it is spread over such a large region (from 15 to 45 km.) that the average molecular density is very little affected by it.) It is thus highly improbable that the increased velocity of sound in the upper regions should be due to a decrease of m. There is a possibility of the increase of velocity due to an increase of γ. Since, however, the constituents and the proportions are same up to 80 km., the value of γ must necessarily be same throughout.

Temperature inversion.—It may, therefore, be concluded that the only possible explanation of the existence of zones of abnormal audibility is that there is a temperature inversion somewhere above the stratosphere. Such a hypothesis of temperature inversion has been put forward from time to time by various investigators [1]; but it is due to F. J. W. Whipple of England [2] that a complete explanation of the phenomenon has been developed on the basis of this hypothesis.

The most plausible cause of temperature inversion is absorption of solar ultraviolet rays by the ozone in the middle atmosphere. Gowan [3], by assuming the ozone distribution as that obtained by the 'Umkehr' method [Chapter IV], has calculated the probable temperature rise in the middle atmosphere when the ozone is mixed with water vapour in various proportions. His calculations show that, in general, there is a rise of temperature due to absorption of solar rays in the region where sound waves are reflected. These points are discussed in detail in Chapter IV.

2. PROPAGATION OF ENERGY OF SOUND WAVES IN THE UPPER ATMOSPHERE

Two questions naturally arise in connection with the propagation of sound waves in the upper atmosphere:

(i) How will the propagation of energy be affected when the medium becomes more and more rarefied? In other words, how will the amplitude and pressure of sound wave vary as it proceeds to higher and higher regions?

(ii) How will the direction of the rays be modified by variations of air currents and temperature?

In regard to the propagation of energy we may distinguish two cases: (*a*) The case of propagation in the stratosphere where the density decreases

exponentially but the temperature remains constant, and (*b*) the case of propagation in the troposphere where, besides the decrease of density, the *temperature also decreases linearly with height.*

(*a*) Propagation in the stratosphere

In a medium of density ρ (or pressure p) and at temperature T the velocity of sound is given by the well-known relation

$$v=\sqrt{\frac{\gamma kT}{m}}=\sqrt{\frac{\gamma p}{\rho}}, \qquad \dots \quad \dots \quad \dots \quad (1)$$

where k, the Boltzmann constant, is $1{\cdot}372\times10^{-16}$ erg. deg.$^{-1}$, m is the mean molecular mass of the constituents of the medium and γ is the ratio of specific heats for air.

The distributions of pressure and density in an atmosphere which is in isothermal equilibrium are given by the relations

$$p=p_0e^{-\frac{mg}{kT}h}=p_0e^{-\frac{g\gamma}{v^2}.h}=p_0e^{-\theta h}, \qquad \dots \quad \dots \quad (2)$$

and

$$\rho=\rho_0e^{-\theta h},$$

where p_0 and ρ_0 denote the values at zero level and θ the barometric constant is equal to $mg/kT=g\gamma/v^2$, v being the velocity of sound in the region.

Consider the propagation of a plane wave in such an atmosphere, vertically upwards or downwards, and suppose the amplitude of vibration to be small. Let h be the height of an elementary volume in the state of rest. At any time t, when the wave is progressing, let its height be ζ; ζ is thus a function of h and t. Similarly, the pressure p and density ρ can be expressed as functions of h and t; $\rho(\xi, t)$ is thus the density of a volume element at height $\zeta(h, t)$ at the time t; it was of density ρ at the time of rest when it was at the position h.

Consider now a small element of density ρ and of unit cross-section and thickness dh lying between h and $h+dh$. This element, when disturbed by the passage of a sound wave, will lie between ζ and $\zeta+(\partial\zeta/\partial h)\,dh$ and its density will be changed to $\bar{\rho}$. Therefore

$$\left(\frac{\bar{\rho}}{\rho}\right)=\left(\frac{\partial\zeta}{\partial h}\right)^{-1}.$$

The force acting on the volume element is $\bar{p}$ in the upward direction and $\bar{p}+(\partial\bar{p}/\partial h)\,dh+\rho g dh$ in the downward direction. The equation of motion is thus

$$\rho\frac{\partial^2\zeta}{\partial t^2}dh=-\frac{\partial\bar{p}}{\partial h}dh-\rho g\,dh$$

or

$$\frac{\partial^2\zeta}{\partial t^2}=-\frac{1}{\rho}\frac{\partial\bar{p}}{\partial h}-g. \qquad \dots \quad \dots \quad \dots \quad (3)$$

Further, the compressions and rarefactions produced by the sound waves are adiabatic; the variations of pressure and density are therefore given by

$$\frac{\bar{p}}{(\bar{\rho})^\gamma} = \frac{p}{(\rho)^\gamma}$$

or

$$\bar{p} = \left(\frac{\bar{\rho}}{\rho}\right)^\gamma p.$$

Substituting the values of $\left(\frac{\bar{\rho}}{\rho}\right)$ and p we have

$$\bar{p} = p_0 e^{-\theta h}\left(\frac{\partial \zeta}{\partial h}\right)^{-\gamma}$$

Differentiating

$$\frac{\partial \bar{p}}{\partial h} = -\left\{\theta p_0 e^{-\theta h}\left(\frac{\partial \zeta}{\partial h}\right)^{-\gamma} + \gamma p_0 e^{-\theta h}\left(\frac{\partial \zeta}{\partial h}\right)^{-(\gamma+1)} \frac{\partial^2 \zeta}{\partial h^2}\right\}.$$

Substituting the values of $\frac{\partial \bar{p}}{\partial h}$ in Eq. (3) and also the values of θ and v from Eqs. (1) and (2) we get the differential equation for wave motion

$$\frac{\partial^2 \zeta}{\partial t^2} - v^2\left(\frac{\partial \zeta}{\partial h}\right)^{-(\gamma+1)} \frac{\partial^2 \zeta}{\partial h^2} + g\left\{1 - \left(\frac{\partial \zeta}{\partial h}\right)^{-\gamma}\right\} = 0. \qquad .. \qquad (4)$$

For vibration of small amplitudes

$$\zeta(h, t) = h + \xi(h, t).$$

Eq. (4) therefore becomes

$$\frac{\partial^2 \xi}{\partial t^2} - v^2 \frac{\partial^2 \xi}{\partial h^2} + \gamma g \frac{\partial \xi}{\partial h} = 0. \qquad .. \qquad .. \qquad .. \qquad (5)$$

Assuming ξ to be simple harmonic

$$\xi \propto e^{j\nu t}, \text{ where } j = \sqrt{-1},$$

we obtain

$$v^2 \frac{\partial^2 \xi}{\partial h^2} - \gamma g \frac{\partial \xi}{\partial h} + \nu^2 \xi = 0,$$

or

$$\frac{\partial^2 \xi}{\partial h^2} - \theta \frac{\partial \xi}{\partial h} + \frac{4\pi^2}{\lambda^2} \xi = 0, \text{ since } \nu = \frac{2\pi v}{\lambda},$$

or

$$\frac{\partial^2 \xi}{\partial h^2} - \theta \frac{\partial \xi}{\partial h} + K_s^2 \xi = 0, \text{ where } K_s = \frac{2\pi}{\lambda}.$$

To solve this equation, put $\xi = e^{nh}$. The value of n is then given by

$$n = (\theta \pm \sqrt{\theta^2 - 4K_s^2})/2.$$

Since $\theta \ll K_s$, $n = \frac{\theta}{2} \pm jK_s\left(1 - \frac{\theta^2}{8K_s^2}\right)$,

so that

$$\xi = C \cdot e^{\frac{h\theta}{2}} \cdot e^{j\nu t} e^{\pm jK_s'h}.$$

$$\text{Real part of } \xi = C\, e^{\frac{\theta h}{2}} \cos (\nu t \pm K'_s h) \quad \ldots \quad \ldots \quad \ldots \quad (6)$$

where C is a constant,

$$\text{and } K_s' = K_s\left(1 - \frac{\theta^2}{8K_s^2}\right).$$

From Eq. (6) it is seen that *the amplitude of vibration of air particles increases with height.* It is also seen that the wave velocity is not quite independent of wavelength. There is slight dispersion in the isothermal region.

Now, the flow of energy E per second per unit area is the product of the velocity of sound v and the average energy density over a complete period and is given by

$$E = \tfrac{1}{2} v\rho \xi^2_{\max} \nu^2.$$

Since $\xi^2_{\max} = \xi^2_{0\max}\, e^{\theta h}$, where ξ_0 is the value of ξ at zero-level, $\rho = \rho_0 e^{-\theta h}$ and v at any height in the stratosphere is approximately equal to v_0 at the base, we have

$$E = \tfrac{1}{2} v_0 \rho_0 \xi^2_{0\max} \nu^2.$$

Thus the rate of energy propagation remains constant throughout and is equal to that at the base.

(*b*) Propagation in the troposphere

The troposphere is approximately in adiabatic equilibrium. We have from Chapter I, Sec. 3(*c*), the relation for such an atmosphere

$$\frac{A\gamma}{\gamma - 1}\left(\rho^{\gamma-1} - \rho_0^{\gamma-1}\right) = -gh.$$

Or, since the temperature at any height is equal to $\frac{Am}{k}\rho^{\gamma-1}$,

$$T = T_0 - \frac{(\gamma-1)m}{\gamma k} gh.$$

Hence,

$$T = T_0\left(1 - \frac{\gamma-1}{v_0^2} gh\right),$$

where $v_0 = \sqrt{\frac{\gamma k T_0}{m}}$, the velocity of sound at the surface of the earth.

Putting $\beta = \frac{\gamma - 1}{v_0^2} g$,

$$T = T_0(1 - \beta h).$$

Further, we have the relation

$$\frac{\rho}{\rho_0} = \left(\frac{T}{T_0}\right)^{\frac{1}{\gamma-1}} = (1 - \beta h)^{\frac{1}{\gamma-1}},$$

where ρ and ρ_0 are the densities at the heights where the temperatures are T and T_0 respectively.

The equation of wave motion in the troposphere can be obtained from Eq. (5) as

$$\frac{\partial^2\xi}{\partial t^2} = v_0{}^2(1-\beta h)\frac{\partial^2\xi}{\partial h^2} - \gamma g\frac{\partial\xi}{\partial h}. \quad \ldots \quad \ldots \quad (7)$$

Assuming $\xi \propto e^{j\nu t}$ as before, we have

$$v_0{}^2(1-\beta h)\frac{\partial^2\xi}{\partial h^2} - \gamma g\frac{\partial\xi}{\partial h} + \nu^2\xi = 0. \quad \ldots \quad \ldots \quad (8)$$

To solve this equation, we write it out in the standard form. By making the following substitutions in Eq. (8),

$\tau = (1-\beta h)$,

$\delta = \gamma/(\gamma-1)$,

$k = 2\pi/\lambda_0$; λ_0 being wavelength at the surface of the earth,

$n = \delta - 1$,

$k_1 = k/\beta$,

we get after simplification

$$\frac{\partial^2\xi}{\partial\tau^2} + \frac{n+1}{\tau}\frac{\partial\xi}{\partial\tau} + \frac{k_1{}^2\xi}{\tau} = 0. \quad \ldots \quad \ldots \quad \ldots \quad (9)$$

Putting $\xi = \omega\tau^{-n/2}$, ω being a function of τ and $\chi = 2k_1\sqrt{\tau}$, Eq. (9) takes the well-known form of Bessel's equation:

$$\frac{\partial^2\omega}{\partial\chi^2} + \frac{1}{\chi}\frac{\partial\omega}{\partial\chi} + \left(1 - \frac{n^2}{\chi^2}\right)\omega = 0.$$

For large values of χ the solution is given by

$$\omega = CQe^{j\chi}/\sqrt{\chi},$$

where C is a constant,

$$\text{and } Q = \left\{1 + \frac{1-4n^2}{1\,.\,(8j\chi)} + \frac{(1-4n^2)(3^2-4n^2)}{1\,.\,2\,.\,(8j\chi^2)} + \ldots\right\}.$$

Or $Q \fallingdotseq 1$, if the higher terms of the series which are small compared to unity are neglected.

Hence

$$\xi = C\,.\,\frac{e^{j\{\nu t + 2k_1(1-\beta h)^{\frac{1}{2}}\}}}{\sqrt{2k_1}\,(1-\beta h)^{\frac{1}{2}(n+\frac{1}{2})}}.$$

$$\text{Real part of } \xi = \frac{C\left(\dfrac{\beta}{2k}\right)^{\frac{1}{2}}\cos\left\{\nu t + \dfrac{2k}{\beta}(1-\beta h)^{\frac{1}{2}}\right\}}{(1-\beta h)^{\frac{3}{2}}}, \quad \ldots \quad (10)$$

since $(n+\frac{1}{2})$ is approximately equal to 3.

From (10) we find that *the amplitude of vibration of air particles*, as in the case of the stratosphere, *increases as h increases*, i.e., as we ascend higher up in the atmosphere.

We can now calculate the amount of energy E flowing per second per unit area at a height h.

We have $$E = \tfrac{1}{2}v\rho\xi^2_{\max}\nu^2.$$

Making the substitutions

$$v = v_0(1-\beta h)^{\frac{1}{2}};\ \rho = \rho_0(1-\beta h)^{\frac{1}{\gamma-1}} \text{ and } \xi_{\max} = \frac{\xi_{0\max}}{(1-\beta h)^{\frac{1}{2}}},$$

where ξ_0 is the value of ξ at the surface of the earth, we obtain

$$E = \tfrac{1}{2}v_0\rho_0\xi^2_{0\max}\nu^2\,[1-\beta h]^{\left(\frac{1}{\gamma-1}-\frac{5}{2}\right)}.$$

Substituting the value of γ we finally get $E = \tfrac{1}{2}v_0\rho_0\xi^2_{0\max}\nu^2$ which is equal to the amount of energy flowing per second across unit area at the surface of the earth ($h = 0$).

The preceding analysis shows that though the atmosphere becomes rarer as one ascends in the troposphere or in the stratosphere, the amount of energy propagated by a plane sound wave remains constant throughout and is equal to that at the base of the atmosphere. This is because the effect of decrease of density upwards is exactly compensated by the increase of amplitude of oscillations. The propagation of energy thus proceeds exactly in the same way as in the case of a homogeneous medium.

(*c*) Limiting height of sound wave propagation: Damping due to viscosity and heat conduction

The remarks made in the last paragraph of the preceding section seem to imply that plane sound waves would proceed with undiminished intensity (i.e., carry constant amount of energy) to any height of the atmosphere, howsoever rarefied it might be. Actually, however, this is not so. The waves, as they proceed upwards, are weakened gradually by the damping actions of viscosity and heat conduction, the effect due to the latter increasing rapidly as the length of the mean free path of the air particles approaches the length of the sound waves. On account of this, waves of shorter length are damped out at much lower heights than waves of greater length. We now proceed to discuss these effects after Schrödinger [4].

Up to the tropopause, the mean free path of the molecules is comparatively small so that the damping effect is negligible. But, at heights much above the troposphere this is no longer the case. Consider the region where the wavelength is of the same order as the mean free path (the maximum wavelength of sound is about 30 m., and the mean free path of air molecules at height 140 km. above ground is of the order of 10 m.). In such a region, since the thermal velocity of the molecules is of the same order as the velocity of sound waves, considerable number of molecules will move out, during the period of one oscillation from the region of compression to that of rarefaction (and *vice versa*) between which there are large

temperature and pressure differences. The effect of motion of these molecules will therefore be to equalize the temperature and the pressure differences and to effectively check the progress of the wave by dissipating its energy to thermal energy. This effect increases as the sound waves proceed to the more rarefied part of the atmosphere. If K be the damping in traversing the path h, then according to Schrödinger K is given by

$$K = -\left[\frac{4\pi^2(\gamma-1)}{\gamma}q+\frac{16\pi^2\eta}{3\rho}\right]\frac{h}{\lambda^2 v} \quad .. \quad .. \quad (11)$$

where q—thermometric conductivity,
η—coefficient of viscosity,
ρ—density of air,
v—sound velocity,
λ—wavelength of sound.

The experimental values at S.T.P. are

$q = 0{\cdot}505\ Cl$

$\eta/\rho = 0{\cdot}285\ Cl$

where C—root mean square velocity of molecules $= v\sqrt{3/\gamma}$
and l—mean free path of the molecules
$= l_0 e^{\theta h}$, l_0 ($= 10^{-5}$ cm.) being the mean free path at atmospheric pressure and $\theta = mg/kT$.

Hence by substituting these values in (11) we get

$$K = -30{\cdot}1\frac{lh}{\lambda^2} \quad .. \quad .. \quad .. \quad .. \quad (12)$$

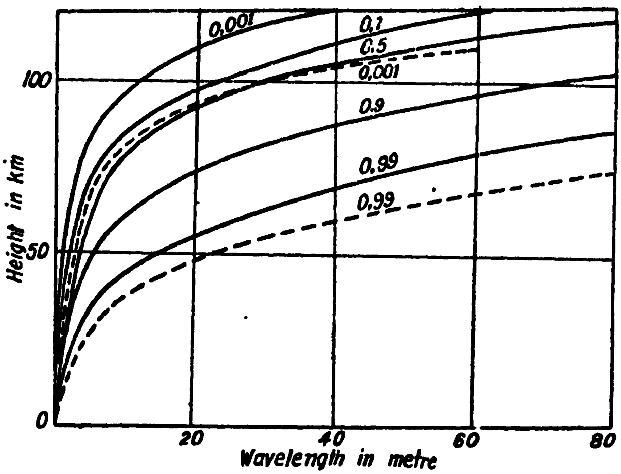


FIG. 2. Illustrating the heights above the ground at which sound waves of different lengths suffer a fall of intensity by exp K (= 0·99, 0·9, 0·5, etc.) in traversing a length of path equal to 1 km. The curves show that longer waves can go higher up than shorter ones. The crowding of the curves in the upper part of the figure indicates that once damping becomes marked, it increases rapidly with height. The full line curves are for the temperature 0°C. The two broken line curves are for a lower temperature, − 45°C. (After Schrödinger.)

If we put $h = 10^5$ cm. $= 1$ km., then from (12) we get the height h_1 above ground at which waves of length λ suffer a fall of intensity by e^K in traversing a path length 1 km.

$$\frac{1}{\theta}\left\{2\log\lambda + \log\frac{-\log e^{K}}{30\cdot 1}\right\}.$$

Putting for e^K the values 0·99, 0·9, 0·5, etc., Schrödinger obtained the curves given in Fig. 2. An inspection of the curves shows that longer waves can go higher up than the shorter ones. At high altitudes of the order of 80 km. even these longer waves are attenuated greatly. If the temperature is lower, the attention is still greater.

3. PATH OF A SOUND RAY IN THE ATMOSPHERE

(a) Effect of temperature gradient

The path of a sound ray can now be traced when, as a result of temperature gradient, the refractive index changes from one stratum to another.

Consider the simple case when there is no horizontal gradient of temperature. This, though not strictly true in practice, is approximately so; the vertical gradient is always much steeper than the horizontal one, even when such a gradient exists. The sound path under this simplifying assumption is representable by a curve lying in vertical plane instead of by a more complicated space curve. If the atmosphere is divided into isothermal layers, the sound velocity in a particular layer may be taken as constant. Let v_1, v_2, v_3, etc. be the velocities and i_1, i_2, i_3, etc. the angles of incidence and refraction as shown in Fig. 3. From the laws of refraction

$$\frac{v_1}{v_2} :: \frac{\sin i_1}{\sin i_2}; \quad \frac{v_2}{v_3} :: \frac{\sin i_2}{\sin i_3}.$$

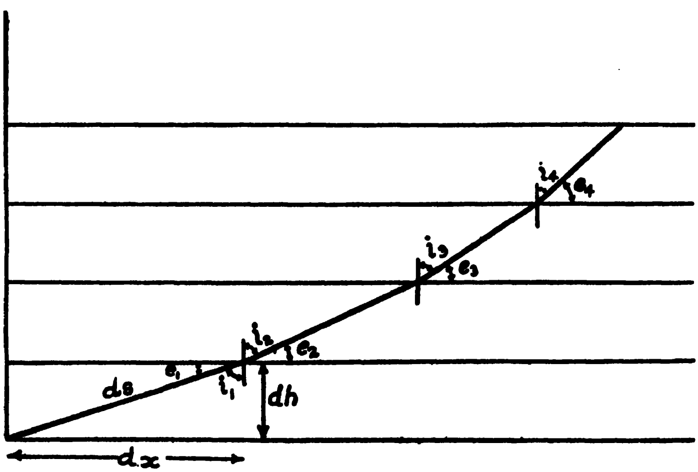


FIG. 3. Illustrating how the path of a sound ray can be traced when, as a result of temperature gradient, the refractive index changes from one stratum to another.

These equations can also be written as

$$\frac{v_1}{\sin i_1} = \frac{v_2}{\sin i_2} = \text{constant},$$

or,
$$\frac{v_1}{\cos e_1} = \frac{v_2}{\cos e_2} = \text{constant},$$

when expressed in terms of the emergence angles, complementary to i_1 i_2, etc.

In the troposphere there is a fall of temperature with height and, as a consequence, the velocity falls continuously as the wave ascends. All rays, with the exception of the strictly vertical ones, follow therefore upwardly curved paths.

Consider now a region where the temperature instead of decreasing increases upwards. In such a region the ray is more and more bent away from the normal so that the angle of incidence is gradually increased until the ray becomes incident on a layer at the critical angle. The phenomenon of total reflection then occurs and the ray follows an exactly symmetrical path downward and reaches the earth. The fact that sound reflections occur from higher level shows that there might be such a region of positive temperature gradient above the stratosphere—the region of constant temperature.

The total horizontal distance traversed by the ray, the time taken to travel from the source to the point of observation and the temperature at the highest point of the trajectory in a region of rising temperature are easily obtained as follows.

For an element of ray path ds (Fig. 3),

$$\frac{dh}{ds} = \sin e, \quad \frac{dx}{ds} = \cos e;$$

hence
$$\frac{dx}{dh} = \cot e.$$

Therefore
$$x = \int_0^{h_{max}} \cot e \, dh$$

and the total horizontal distance $= 2 \int_0^{h_{max}} \cot e \, dh$, ... (13)

where h_{max} represents the height reached by the ray.

The time taken by the sound to travel from the source to the point of observation is given by

$$\tau = 2 \int_0^{h_{max}} ds/v \qquad \text{.. .. .. .. (14)}$$

$$= 2 \int_0^{h_{max}} dh/(V \cos e \sin e), \text{ where } V = \frac{v}{\cos e}.$$

At the highest point of the trajectory $e = 0$ so that the velocity is in the horizontal direction. V is therefore the velocity at this point and is called the *characteristic velocity*.

Now, the velocity at any height is related to the temperature at that height by

$$v = 20{\cdot}1\sqrt{T}\text{ m./sec.} \quad .. \quad .. \quad .. \quad (15)$$

A knowledge of the characteristic velocity therefore enables one to estimate the temperature at the highest point of the trajectory.

The above relation connecting v and T is obtained from Eq. (1) by substituting the values $\gamma = 1{\cdot}4$, $k = 1{\cdot}37 \times 10^{-16}$ erg. deg.$^{-1}$ and $m = 4{\cdot}7 \times 10^{-23}$ gm. For use of this expression at any height h, it is of course assumed that γ and m are constant with height.

(*b*) Effect of wind

The presence of air current influences the path of the sound ray. If the velocity of the air current, which we assume to be in the horizontal direction, is constant with height, the effect is rather small. But if the velocity increases with height, the influence becomes appreciable. If a sound ray proceeds obliquely upwards, it experiences higher and higher wind velocity as it ascends. The effective horizontal velocity of the sound thus increases and the direction of the ray deviates more and more towards the ground. It ultimately becomes horizontal and, as a result of the gradient of the wind velocity on the wave front, is bent back towards the earth. If, however, the sound ray has a component in a direction opposite to that of the wind, the ray as it ascends becomes more and more steeply inclined to the vertical and cannot return to the earth. The sound is thus better heard in the direction of the wind than against it.

A simplified mathematical analysis of the wind effect, after Emden [5], is given below. It is assumed that the temperature decreases and the wind velocity increases—the latter linearly—with height.

In Fig. 4, OM is an element of wave-front at time t. Let v be velocity of sound and w that of wind. In unit time the middle point M moves to M' due to sound velocity and at the same time is carried by wind velocity to M''. The resultant displacement is MM'' and the ray velocity is v'. The new wave-front is given by $O''M''$. The velocity v_n along the wave-normal MM' is given by

$$v_n = v + w \cos e_n,$$

where e_n is the angle between v and the horizontal x-direction.

It can be shown that $\dfrac{v_n}{\cos e_n} = \text{constant}$. Hence,

$$\frac{v}{\cos e_n} + w = \frac{v_0}{\cos e_0} + w_0,$$

v_0, e_0, w_0 being the values corresponding to zero-level.

Again, the ray velocity

$$v' = \sqrt{v^2+w^2+2vw\cos e_n}$$

$$= w\cos e_s + \sqrt{v^2-w^2\sin^2 e_s},$$

where e_s is the angle between v' and the horizontal direction.

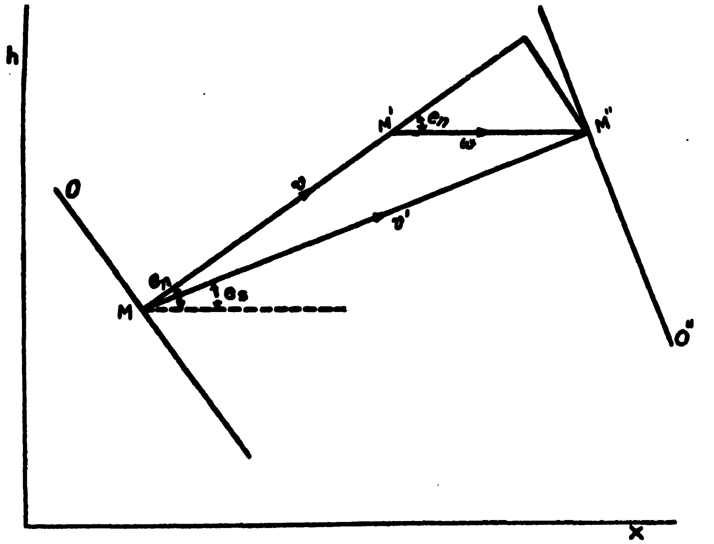


FIG. 4. Tilting of wave-front of sound due to the effect of wind. *OM* is the original wave-front and *O''M''* is the tilted wave-front due to the effect of wind velocity.

The temperature decreases linearly with height. Hence, strictly speaking, the velocity decreases as the square root of height. However, within the small range of height within which the bending of the sound ray takes place, the velocity may be assumed to decrease linearly. Further, since the wind velocity increases linearly, we may put

$$v = v_0(1-ah)$$

and $$w = w_0(1+bh).$$

Also, omitting the suffix n, we have

$$\cos e = \frac{v}{\dfrac{v_0}{\cos e_0}+w_0-w}$$

$$= \frac{\cos e_0}{1+Bh}, \quad \text{where } B = a - \frac{w_0 b}{v_0}\cos e_0.$$

Considering the ray we have

$$dx = (v\cos e + w)dt$$

and $$dh = v\sin e\, dt.$$

Therefore,

$$\frac{dx}{dh} = \frac{v\cos e + w}{v\sin e} = \frac{\cos e_0 + \dfrac{w}{v}(1+Bh)}{[(1+Bh)^2 - \cos^2 e_0]^{\frac{1}{2}}}.$$

This, on simplification and integration, yields

$$x = \frac{\cos e_0}{B}\cdot\left(\sqrt{2Bh+\sin^2 e_0} - \sin e_0\right)$$
$$+\frac{w_0}{v_0}\left\{1 - w_0(a+b)\,(\sin^2 e_0 - Bh\cos^2 e_0)/(3Bv_0\cos^2 e_0)\right\}$$
$$\left\{\left(\sqrt{2Bh\cos^2 e_0 + \sin^2 e_0} - \sin e_0\right)/(B\cos^2 e_0)\right\}.$$

Taking the particular case when the ray starts off in a horizontal direction so that $e_0 = 0$, we have

$$x = \left[1 + \frac{w_0}{v_0}\left\{1 + \frac{(a+b)hw_0}{3v_0}\right\}\right]\sqrt{2\sigma_1 h},$$

where $$\sigma_1 = \frac{1}{a - \dfrac{w_0 b}{v_0}}.$$

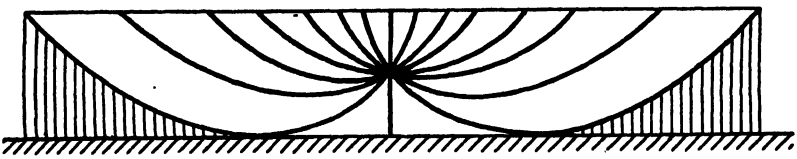


FIG. 5. Paths of sound rays when the source is situated at a height above the ground and the temperature decreases upwards. The regions shaded by vertical lines are regions of silence. (After Emden.)

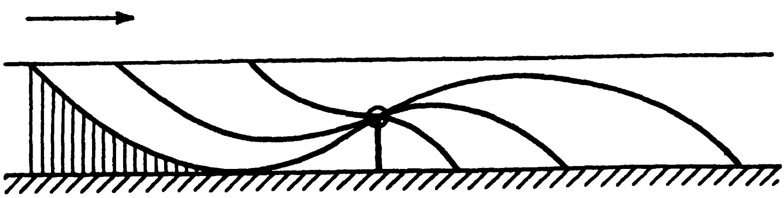


FIG. 6. Effect of wind (indicated by the arrow) on the state of affairs depicted in Fig. 5. (After Emden.)

Figs. 5, 6, 7 depict, after Emden [5], the paths of the sound rays for some typical cases. Fig. 5 shows the case when the sound source is situated at a height and the temperature decreases upwards. Fig. 6 depicts the case when there is an air current over and above the temperature gradient as in Fig. 5. It is clearly seen that audibility is greater in the direction of the wind than against it. Fig. 7 is drawn for a specific case. The source is

situated on the surface of the earth; the atmosphere is still up to 0·37 km. above which there is a gradient of 4 m. per sec. per km. of ascent. The tem-

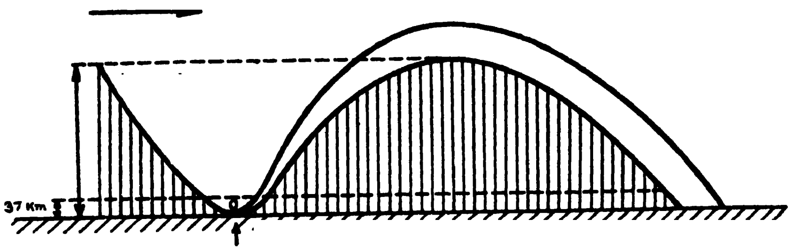


FIG. 7. Paths of sound ray for a specific case: The temperature near the ground, 294°K (v_0 = 342 m./sec.); temperature gradient,—6·2°K per km.; no wind up to a height of 0·37 km., above which there is a gradient of 4 m./sec. per km. of ascent. The regions shaded by vertical lines are the regions of silence. (After Emden.)

perature gradient is −6·2°K per km. and the temperature near the ground is taken to be 294°K (hence $v_0 = 342$ m./sec.). The shaded zones represent the regions where no sound is heard.

4. EXPERIMENTAL STUDY OF ANOMALOUS SOUND PROPAGATION—TEMPERATURE DISTRIBUTION IN THE MIDDLE ATMOSPHERE

From what has been said above it is obvious that the heights of the trajectories of the anomalously propagated waves and the corresponding characteristic velocities enable one to determine the temperature distribution in the hot region of the upper atmosphere. We shall presently describe methods for measuring and calculating these quantities. The experimental data necessary for these calculations are the times taken by the anomalously propagated sound waves to travel from the source to a number of receiving stations suitably distributed. For systematic collection of the data it is not sufficient to depend on aural evidence or to wait for accidental explosions. Artificial explosions are arranged for making measurements at different hours of the day or night, in different seasons and also at different latitudes. An essential apparatus for making such measurements is some delicate device for recording the arrival of sound pulses. Before we proceed to discuss the methods of estimating the heights of the trajectories we shall briefly describe the working principles of three of the devices which have been developed for the purpose.

(a) Recording instruments

Undograph.—This recorder, designed by Kühl, has been widely used in Germany [6]. Its essential component, the deflecting element, is shown in Fig. 8. A rectangular membrane (length $2b$, breadth $2a$, thickness d and

made of material of density σ) is supported by wire XY passing through an axis of symmetry. The wire is fixed at its two ends with a certain amount of tension. One-half of the membrane lying on one side of the axis is made 'dead' by being shielded from the incoming wave while the other half is exposed to same. A small mirror is fixed to the wire, so that rotation of the membrane is registered by the movement of a light-pointer.

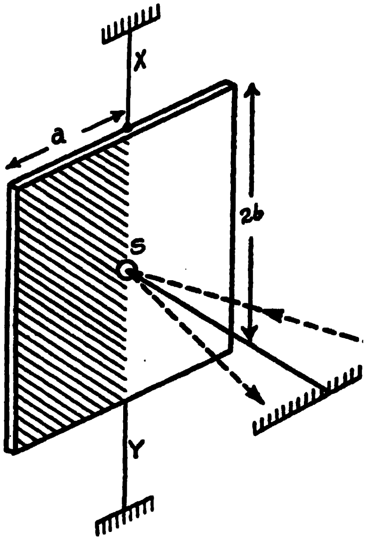


FIG. 8. Illustrating the deflecting element, a thin membrane, of the Undograph. The shaded portion of the element is made dead by shielding it from the incoming wave. The rotation of the mirror S is registered by the movement of the light pointer.

If a sound wave incident on the exposed half of the wing produces a change of pressure δP then the deflection D of the light-pointer is given by

$$D = 2A\frac{a^2b}{\tau}\,\delta P,$$

where A is the distance between the scale and the mirror S, and τ the torsional constant of the wire.

The free period of oscillation of the system is given by

$$T = 4\pi\sqrt{\frac{a^3b\sigma d}{3\tau}\left\{1+\left(\frac{d}{2a}\right)^2\right\}},$$

assuming, of course, that the axis is vertical. Combining the above two equations we get

$$D = \frac{3}{8\pi^2}A\delta P\frac{T^2}{\sigma a d\left(1+\frac{d^2}{4a^2}\right)}$$

$$\doteqdot \frac{3}{8\pi^2}A\delta P\frac{T^2}{\sigma a d}.$$

The above relation shows that sensitivity is increased by decreasing the dimensions of the wing so that if the dead part of the wing is removed the equation becomes

$$D = \frac{3}{2\pi^2} A\delta P \frac{T^2}{\sigma a d}.$$

The equation further shows that an increase in the transverse axis of the wing will not contribute to the sensitivity of the instrument; for, the advantage gained by a greater air surface is compensated by the increase in the moment of inertia.

Hot-wire microphone.—This instrument developed by Tucker and Paris [7] consists essentially of a closed vessel in the form of a Helmholtz resonator, across the opening of which is stretched a thin wire which may be heated electrically. When sound waves are incident on the orifice of the resonator the volume of air inside begins to vibrate and a portion of the air flows in and out through the orifice. It thus cools and changes the resistance of the hot-wire. The variation of the resistance then indicates the arrival of sound waves. The resonator generally has a volume varying from 300 c.c. to 50 c.c. and has a sharp resonance characteristic. The neck of the resonator, ending in the orifice, has ordinarily a length of 2 cm. and diameter 1 cm. The length of the neck cannot be decreased below a limit of about 1 cm. because the hot-wire is then exposed to transient air currents. Resonance adjustment may be made by increasing or decreasing the volume of the resonator and/or the length of the neck.

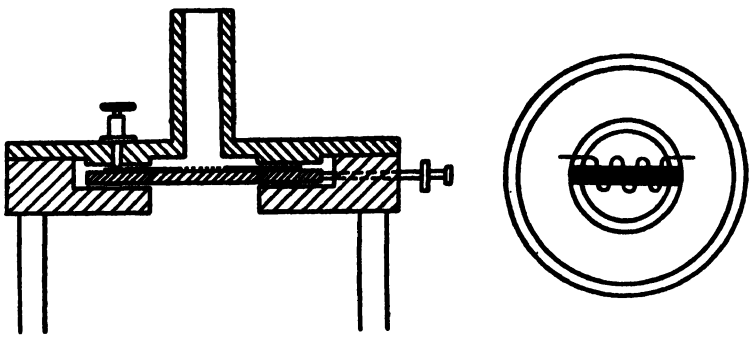


FIG. 9. Illustrating the principle of Tucker and Paris' hot-wire microphone. The heating element is shown on the right fitted on a disc with a central hole across which there is a support for the wire (shown in black). The disc is mounted at the mouth of a Helmholtz resonator (left diagram). The resonator ends in a neck with a diameter of about 1 cm. and a communicating tube with the outer air of length about 2 cm. The tube protects the hot-wire element from transient air currents outside.

The heating element is made of very thin platinum wire (dia. 0·0006 cm.) bent into loops as shown in Fig. 9 and supported by a rod across a small hole (0·6 cm. diameter) made in a circular mica disc. The system is suitably mounted on the orifice and the wire is heated just below red heat by passing a current (about 30 mA) so that the resistance of the wire increases from

about 150 ohms to about 350 ohms. The resistance is now balanced in a bridge circuit containing a sensitive Einthoven galvanometer. Any change of resistance due to the incidence of sound waves disturbs the balance of the bridge and produces a deflection in the galvanometer.

A great advantage of this instrument is that it is free from the effect of spurious vibration such as may be communicated through the mounting, because, the hot-wire is sensitive only to the air vibration. Another advantage of the hot-wire microphone over the Undograph is that the recording of sounds received at different stations may be made on one and the same film. A comparison of the times of arrival of sound can therefore be made with greater precision.

Electromagnetic micro-barograph.—This instrument, designed by Benioff [8] for recording sudden and slight changes in air pressure for purposes of seismographic investigations, may be used for studying anomalous propagation of sound. The instrument consists essentially of a moving coil loud-speaker (permanent magnet type) mounted on one side of a closed chamber of capacity about $\frac{1}{8}$ cu.m. The effective diameter of the loud-speaker cone is about 15 cm. Since the natural frequency of vibration of its cone assembly is about 150 cycles per second, the displacement of the cone for frequencies less than 50 cycles is approximately proportional to the increment of pressure. The induced e.m.f. in the coil is thus proportional to the rate of change of pressure. The output current passes to a galvanometer, the deflections of which can be recorded in the usual way. The sensitivity of the instrument is such that a pressure change of 1 dyne/cm.^{2} produces a deflection of 1 mm. in a galvanometer of period 1 sec.

(b) Temperature distribution: Height of the trajectory

As already mentioned, the temperature in the hot region of the atmosphere can be calculated from the value of the characteristic velocity, i.e., the velocity of sound at the highest point of the trajectory of the anomalously propagated sound wave. This latter can be determined either from the measurement of the angle of descent of the downcoming wave or from the so-called travel-time curve, i.e., a curve showing the relation between the time of travel and the horizontal distance traversed by the wave. The two methods are discussed below.

(i) *From angle of descent.*—The angle of descent may be measured after Meisser by recording the exact hours of arrival of the sound waves at three different stations, a few hundred metres apart, situated conveniently at three corners of a triangle of approximately equilateral shape [9].

Let P_0, P_1 and P_2 be the three receiving stations. Imagine a rectangular co-ordinate system with origin at P_0, such that the x-axis is along the line joining P_0 and P_1 and z-axis vertically upward. The co-ordinates of the three stations are then—assuming that they lie on a horizontal plane—(0, 0, 0), (x_1, 0, 0) and (x_2, y_2, 0) respectively. Let the times of arrival of a plane downcoming wave at the three stations be

T_0, T_1 and T_2. If the wave-normal makes angles α, β and γ with the three axes and v_0 is the velocity of sound at the ground, then

$$x_1 \cos \alpha = v_0(T_1-T_0) = \Delta_1,$$

$$x_2 \cos \alpha + y_2 \cos \beta = v_0(T_2-T_0) = \Delta_2,$$

and

$$\cos^2\alpha+\cos^2\beta+\cos^2\gamma = 1.$$

From these relations we can find out γ the angle of incidence with the vertical and ϕ the azimuth of the downcoming ray. Thus

$$\sin \gamma = \left[\left(\frac{\Delta_1}{x_1}\right)^2+\left(\frac{\Delta_2}{y_2}-\frac{x_2}{y_2}\cdot\frac{\Delta_1}{x_1}\right)^2\right]^{\frac{1}{2}},$$

$$\tan \phi = \frac{\cos \beta}{\cos \alpha} = \left(\frac{\Delta_2}{y_2}-\frac{x_2}{y_2}\cdot\frac{\Delta_1}{x_1}\right)\bigg/\frac{\Delta_1}{x_1}.$$

The angle of descent is the complement of the angle of incidence. From the angle of descent so obtained, the characteristic velocity, i.e., the velocity at the highest point of the trajectory is obtained at once from the relation

$$V = v_0 \sec e_0, \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (16)$$

where v_0 is the value of velocity of sound at the ground and $e_0 = \pi/2-\gamma$. From the characteristic velocity again, the temperature of the highest point of the trajectory is obtained from Eq. (15), namely,

$$V = 20{\cdot}1\sqrt{T} \text{ m./sec.}$$

F. J. W. Whipple [10] has made a number of measurements of the angles of descent and has calculated therefrom the characteristic velocities. According to him this velocity is about 350 m./sec. for most of the days. This is about 10 m./sec. higher than the velocity of sound near ground at summer temperature of England. As an extreme case Whipple has recorded an angle of descent as great as 35°, corresponding to a characteristic velocity of 420 m./sec. This large value of characteristic velocity may be explained either by assuming a strong air current in the higher regions or by supposing that at some height there exists a very high temperature.

If now the height of the trajectory be known, the distribution of temperature in the hot region can be determined. There is, however, some uncertainty regarding this height and Whipple [10] makes an approximate estimate of the same as follows.

The meteorological data regarding temperature, wind, etc., as available up to the highest point on the day of experiment, are utilized to trace the actual path of the ray from the ground upwards (from the ground the trajectory is assumed to start off in the direction as given by the measured angle of incidence). The trajectory above the highest point so obtained is calculated for two extreme cases of assumed temperature distribution. Firstly, it is assumed that the temperature above is constant up to the highest point of the trajectory. This evidently means that there is a sharp and sudden rise of temperature at the point of reflection. The other

extreme assumption is that there is a uniform rise of temperature from the calculated highest point. The actual path of the ray can now be assumed to lie somewhere between these two extreme cases. Figs. 10 (*a*) and (*b*) illustrate the procedure. The portion *O* to *A* is calculated from the meteorological data. It is assumed that the portion *BZ* corresponds to *OA*. The trajectory for the two extreme cases are *ATB* and *AJB*. For *OATBZ* the calculated travel-time is 786 seconds and is greater than the actual travel-time (see Fig. 10(*b*)). For *OAJBZ* the travel-time is 667 seconds and is less than the observed travel-time. The trajectory *OAPKQBZ* is now drawn by trial corresponding to observed travel-time 720 seconds. For this trajectory the transition from straight to curved path occurs at *P* at height of 33 km. and the height of *K* is 44 km. The height of *k* is thus assumed to be the height of the trajectory. Figs. 10 (*c*) and (*d*) depict the variation of sound velocity and the temperature distribution with height.

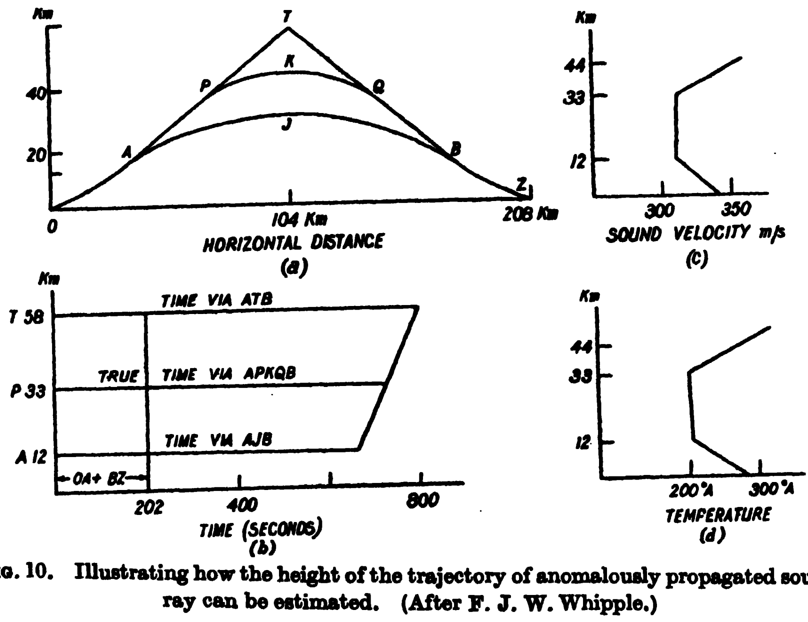


FIG. 10. Illustrating how the height of the trajectory of anomalously propagated sound ray can be estimated. (After F. J. W. Whipple.)

The 'angle of descent' method has also been used by Mathur (under the auspices of the India Meteorological Department) to estimate the temperature and its seasonal variation in the middle atmosphere over the Central Provinces, India [10*a*]. The observations were made during the period May 30, 1946 to Feb. 1, 1947. There were two explosion centres one situated at Ponia (Lat. 23° 40·5′, Long. 80° 23′) and the other at Phulgaon (Lat. 20° 42′, Long. 78° 27′). There were altogether 9 recording stations. Of these, 7 were distributed along the line joining the two centres and 2 along lines nearly at right angles to the same. The most distant recording station was at Kamaredi, 638 km. south of Ponia. The recording apparatus used was one of the Army Sound Ranging Recorders, SR No. 2; MK1

usually employed for localizing enemy mortars. The temperature as measured varied between 324°K at 42 km. to 345°K at 51 km. in summer and between 304°K at 34 km. to 324°K at 54 km. in winter. In the region investigated (40–60 km.) the lapse rate during summer (June) was found to be steeper than during pre-monsoon and winter (October and January). In summer (1946) explosion sounds were better recorded at stations situated south and in winter (1947) pre-monsoon trials were better recorded at stations in the north. These preferred propagations are attributed to upper wind systems.

(ii) *From travel-time curves.*—The times of arrival of the anomalously propagated sound waves at different distances from the source furnish data which, when combined with the knowledge of the actual temperature distribution in the troposphere (obtained directly by balloon sounding, etc.), enable one also to calculate the highest point of the trajectory [11].

The total horizontal distance Δ_s travelled by the ray in the region above the troposphere and the corresponding time t_s are given by

$$\Delta_s = \Delta - D^*$$
$$t_s = t - t^*,$$

where Δ is the horizontal distance between the source and the receiver and t the observed travel-time. D^* is the total horizontal distance traversed by the portions of the ray in the troposphere near the source and near the receiver and t^* the corresponding time.

In the above relations Δ and t are obtained from actual observations and a curve is drawn depicting the relation between the two (Fig. 11). D^* and t^* are evaluated as follows. From Eqs. (13) and (14),

$$D^* = 2\int_0^{h_0} \tan i \, dh$$

$$t^* = 2\int_0^{h_0} \frac{dh}{v \cos i},$$

where h_0 is the height of the tropopause. The velocity v_0 at sea-level and v at any height in the troposphere can be evaluated from a knowledge of the temperature at sea-level and from the distribution of the temperature with height in the troposphere (Eq. 15). The angle i_0 between the ray and the plumb line at sea-level can be obtained from the slope of the $\Delta - t$ curve. Thus,

$$\frac{\sin i_0}{v_0} = \left(\frac{\delta t}{\delta \Delta}\right)_0 . \qquad \ldots \quad \ldots \quad \ldots \quad (17)$$

The angle i at any height is given by

$$\sin i = \frac{v}{v_0} \sin i_0 .$$

Obtaining D^* and t^* in the way indicated above and subtracting from the corresponding values of Δ and t, another curve can be drawn depicting

the relation between Δ_s and t_s, i.e., between distance and time of travel at the top of troposphere (Fig. 12). From this curve the height of the

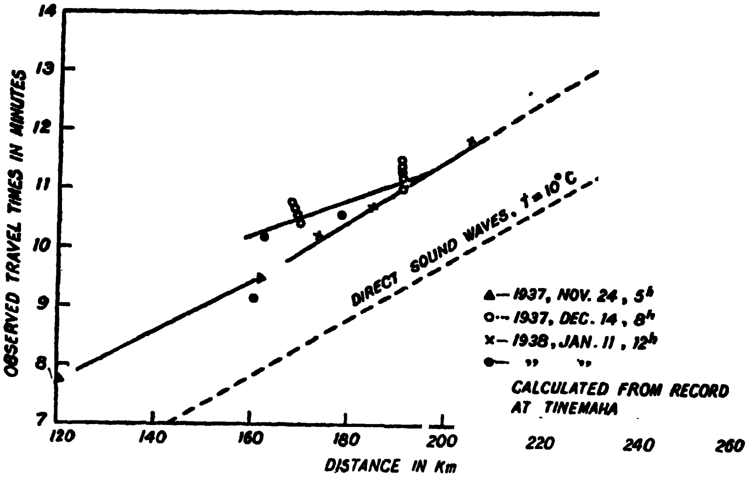


FIG. 11. Experimental travel-time curves for indirect sound wave at Tinemaha, Southern California. (After Gutenberg.)

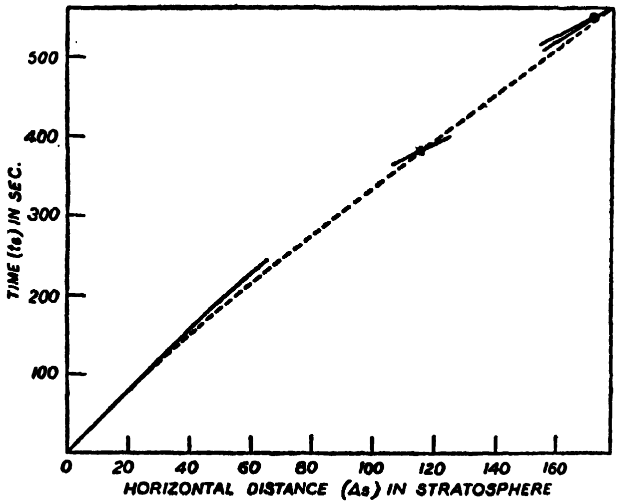


FIG. 12. Calculated travel-time curve for sound waves in the stratosphere corresponding to the data in Fig. 11. (After Gutenberg.)

portion of the trajectory of the sound ray *above the tropopause* can be obtained with the help of the method developed by Herglotz, Wiechert and Bateman for calculating the trajectories in connection with seismological problems. The elevation S of the highest point over the troposphere for a ray covering a horizontal distance $\bar{\Delta}_s$ in the region above the tropopause is given by

$$S = \frac{1}{\pi}\int_0^{\bar{\Delta}_s} \cosh^{-1}\frac{\bar{v}_s}{v_\Delta}\, d\Delta_s, \quad \ldots \quad \ldots \quad \ldots \quad (18)$$

where v_Δ is the value of $\delta\Delta_s/\delta t_s$ at any distance between 0 and $\bar{\Delta}_s$ and $\bar{v}_s$ is the value of the same at the limiting distance $\bar{\Delta}_s$. Since $\delta\Delta_s/\delta t_s$ is obtained from the $\Delta_s - t_s$ curve (Fig. 12), the integration for S can be made graphically. The height h of the point above the surface of the earth is obviously $h_0 + S$.

Since the angle of descent e_0 ($= \pi/2 - i_0$) can be obtained from Eq. (17), the value of the characteristic velocity at height h may be easily calculated from Eq. (16). We thus get the temperature distribution with height (Fig. 13).

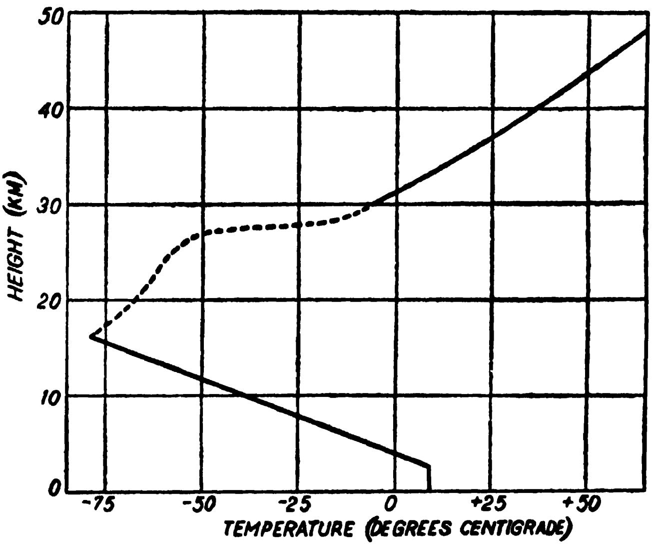


FIG. 13. Temperature distribution in the middle atmosphere in North California in the winter of 1937-38. (After Gutenberg.)

The procedure as given above will be clear from the following example [12]. Referring to Fig. 11 we have the following data :—

Date 1937/38	Δ km.	t m	t s	$(\delta\Delta/\delta t)_0$ m./sec.	v_0 m./sec.	$\sin i_0$	$\sin i_s$	Δ_s km.	t_s m	t_s s
Nov. 24	140	8	36	390	340	·61	·50	118	6	30
Dec. 14	180	10	52	550	340	·44	·36	166	8	58
Jan. 11	200	11	30	370	342	·65	·53	176	9	21

With the values of Δ_s and t_s in the last two columns Fig. 12 is drawn.

From the curve in Fig. 12 the values of S are calculated from Eq. (18). The table below is now prepared giving the distribution of temperature with height, assuming the height of tropopause to be 17 km.

Δ_s km.	t_s sec.	$\bar{v}_s$ m./sec.	S km.	(h_0+S) km.	v m./sec.	Temperature °K	°C
0	0	277	0	17	277	190	−83
50	178	298	..	..	..	..	..
100	337	326	$12\frac{1}{2}$	$29\frac{1}{2}$	326	264	−9
150	485	349	$21\frac{1}{2}$	$38\frac{1}{2}$	349	303	30
200	624	370	$31\frac{1}{2}$	$48\frac{1}{2}$	370	337	64

In the above calculation, it has been assumed that the temperature begins to increase from the level of the tropopause. Actually, however, there is a region of constant temperature just above this level. The sound ray in this region follows a straight course. A correction has therefore to be applied to take account of this region of constant temperature [13].

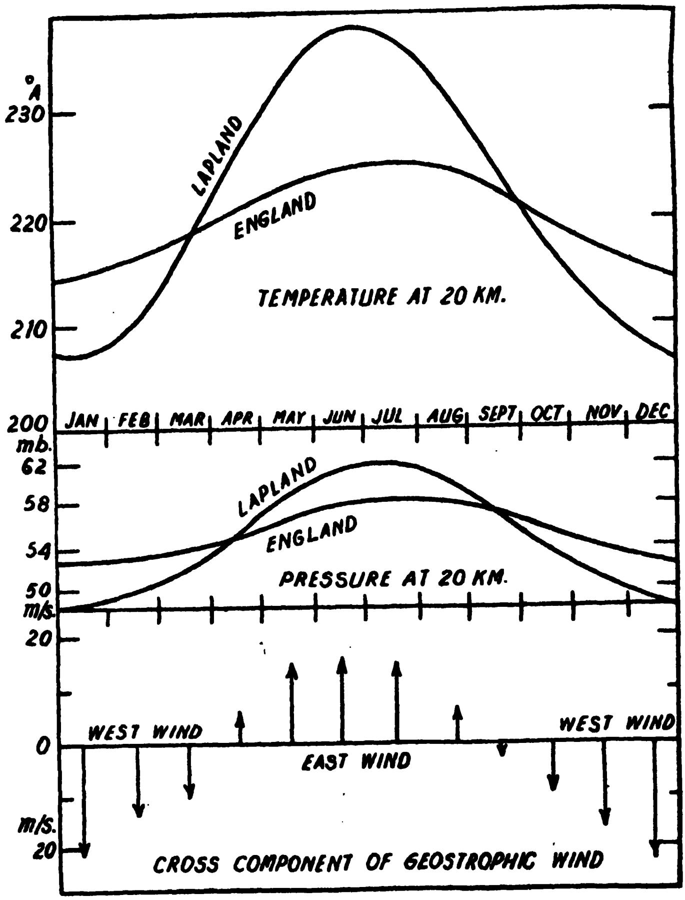


FIG. 14. Illustrating the seasonal variation in temperature, pressure and wind at 20 km. height in England and in Lapland. The wind system in the upper atmosphere produces variation of audibility in different directions of abnormally propagated sound waves. (After F. J. W. Whipple.)

It has been noticed that during different seasons, audibility is favoured in different directions. For instance, it has been found that there is a change for favoured audibility to the west in summer and to the east in winter in England. This is naturally attributed to the changes in direction of the wind at high altitude.

The presence of such winds in the middle atmosphere has been ascribed by F. J. W. Whipple to the presence of pressure difference in high altitudes between places in different latitudes and longitudes. He has studied the records of temperature 20 km. above the ground made by Dines in England and by Rolf in Lapland. The seasonal variations of temperature and pressure at 20 km. level above the ground is shown in Fig. 14. The east wind and the west wind, which will be produced as a result of this difference of pressure during different seasons of the year is shown in the figure. It is to be noted that the transition from west wind to east wind and *vice versa* occurs at the end of March and at the middle of September. These dates agree with the critical dates of abnormal audibility of sound in England.

IIIB—METEORIC PHENOMENA

1. INTRODUCTION

Observation and study of meteoric flashes yield important information concerning temperature and density distribution in upper atmospheric regions. Such studies were first made by Lindemann and Dobson who, by developing a detailed theory of the appearance and disappearance of the meteors, came to the surprising conclusion that the density in the middle atmosphere (50 km. to 60 km.) was much higher than what it used to be believed at the time (1922) on the assumption that the stratosphere and its upper regions were isothermal at the low temperature of 220°K. In fact, the calculated densities could only be reconciled with observations if a high temperature, of the order of 300°K was assumed to prevail in upper stratospheric regions. Other theories and studies of meteor (Sparrow, Öpik, Herlofson, F. L. Whipple) also lead to similar conclusion. These results are therefore in full agreement with that deduced in Chapter IIIA from observation and discussion of the abnormal propagation of sound waves.

In the present chapter we shall first give a short account of the classification and composition of meteors. Methods of observation, theories of meteors and the upper atmospheric temperature and density as may be deduced therefrom will then follow.

It has been found that intense ionization is produced along meteor trails. This ionization contributes to the general ionization of the upper atmosphere produced by ultraviolet solar radiation and by impact of extra-terrestrial particles. Study of meteor-ionization by radio-echo technique

is now a powerful method of determining meteor characteristics and hence of upper atmospheric data.. This will be taken up in Sec. 5.

2. CLASSIFICATION AND COMPOSITION OF METEORS

Meteors are extra-terrestrial particles of mass of a few milligrams which enter the earth's atmosphere and in their rush through it are vaporized by the heat generated and emit light. They are, in fact, by nature projectiles of extremely high speed, 10 to 15 times that of ordinary projectiles. Occasionally the particles may be of much greater size—few hundreds of kilograms when they appear as fire balls and drop on the surface of the earth as meteorites. For example, the Great Greenland Meteor which is the largest preserved meteorite has a mass of $3{\cdot}7 \times 10^4$ kilograms and is now in the American Museum of Natural History. Such a mass, if it were

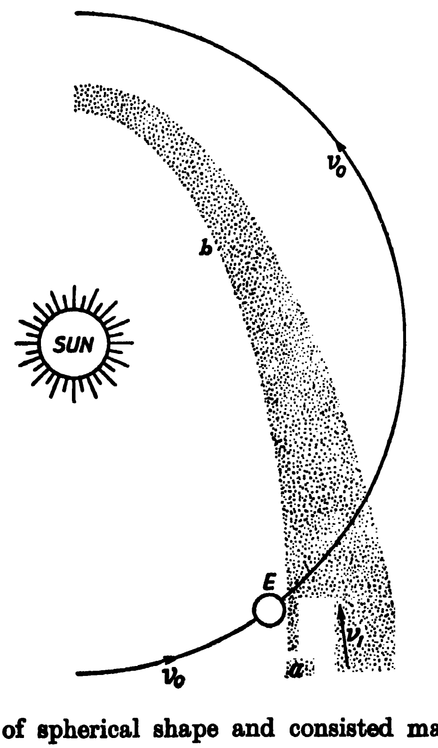


FIG. 1. Origin of shower meteors. The particles in the belt (ab) are debris of defunct comets. The widths of the belts vary greatly and may be estimated from the duration of the shower which is a measure of time the earth takes to cross the belt. Thus, while the Giacobinids shower of October, 1946—a shower of exceptional intensity—lasted only 6 hours, the regularly occurring Perseids shower, though less active, lasts for 21 days. Since the velocity of the earth (v_0) in its orbits is 30 km./sec., the diameter of the cross-section of the swarm (neglecting the fact that the earth's orbit cuts the meteor orbit obliquely) of the former is 400,000 miles, and of the latter 50 million miles. (The earth's and the meteoric particles' orbits are not in the same plane.)

of spherical shape and consisted mainly of iron would be 4·1 metres in diameter. The meteors which disappear in the higher atmosphere are, of course, of much smaller size. Various considerations show that such meteors have the average radius ranging from 0·1 to 0·01 cm.

The meteors may broadly be classified as shower meteors and as non-shower or sporadic meteors.

The origin of shower meteors is currently believed to be defunct or disintegrated comets. The meteoric particles thus belong to different families according to the comet from which they originate and follow

different orbits. Whenever the earth in its annual round crosses the orbit of one such family, the particles belonging to the family are attracted into the atmosphere and meteoric shower belonging to the particular family occurs (Fig. 1).

A list of the major showers together with their dates of maximum, radiants and normal hourly rates is given in Table I. (British Astronomical Association Handbook.) These showers are observable from geographical latitudes $\pm 52°$. (The lists are for meteoric showers as are observed visually. They do not include the 'daylight' showers as are now detected by the radar technique. Reference to daylight meteors will be made in Sec. 5.)

TABLE I

Date of maximum	Shower	Radiant R.A.	Radiant Dec.	Normal hourly rate
January 3 ..	Quadrantids	230°	+ 52°	30 to 40
April 21	Lyrids	270°	+ 33°	7 to 10
August 10–13 ..	Perseids	47°	+ 58°	40 to 60
October 20–23 ..	Orionids	96°	+ 15°	10 to 20
November 16-17 ..	Leonids	152°	+ 22°	10 to 15
December 11–13 ..	Geminids	113°	+ 32°	60

In Table II is given a list of showers of which the parent comets have been identified with good certainty [1].

TABLE II

Shower	Comet	Computed date	Computed radiant
Lyrids ..	1861 I	April 21	271°+34°
June Draconids ..	Pons-Winnecke	June 30	208°+54°
Perseids ..	1862 III	August 11	45°+58°
October Draconids ..	Giacobini-Zinner	October 10	262°+54°
Leonids ..	1861 I	November 15	151°+23°
Andromedids ..	Biela	November 30	23°+44°

The dates of occurrence as given in Table II are computed from the orbits of the original associated comets. It may be noted that for the last two showers the order of the dates of occurrence has been inverted in the course of time. At present Andromedids shower occurs on November 14 and the Leonids on November 16.

Two other identifications may also be mentioned here. F. L. Whipple [2] has shown that Taurids shower has common origin with Encke's comet. Also the Aquarids shower in May is assigned to Halley's comet [3].

For the other major showers which occur regularly the parent comets still remain unknown. It is, of course, possible that they may not have any parent comet at all.

The Giacobinids shower on which many observations have been made, is not included in the lists as it is not regular. It occurred with great intensity on October 9–10, 1946.

The names of the different families, as will be seen from the above tables, are associated with constellations. This is because the observed paths of the meteors belonging to a particular family all meet approximately at a point when projected on the celestial globe. The family to which the shower belongs derives its name from the constellation in the celestial sphere in which this point—called the radiant—is situated. For example, the Leonids is so called because its radiant is in the constellation Leo. The radiants are however not always sharp. The Leonids have a sharp radiant, but the radiants of most of the other annual major showers are scattered over a considerable area. Sometimes the radiants, though scattered over a large area, are concentrated in sharp sub-centres. Such is the case with Orionids. They may also be diffused fairly uniformly over a considerable area as in the case of Perseids. Lack of parallelism between individual meteors clearly mean that though they may be members of the same family and presumably have the same origin they are not moving in identical orbits. As such, at great distances they may not be recognizable as belonging to one family at all. The distribution of meteoric particles being not uniform, there is a main shoal outside which the particles are relatively sparsely distributed. A maximum of meteor shower occurs if the earth in its motion happens to cross the track of a meteor family when the main shoal of the family is near the ecliptic. In the case of Leonids, this maximum occurs at intervals of 33 years.

In the absence of any perturbing action, the meteors of a particular family having a common origin would all follow the same orbit. But, on account of the perturbing action of the planets, particularly that due to Jupiter, the paths of comets and meteor streams are disturbed. It is evident that the amount of perturbation will depend on the time during which the meteoric family or the comet is influenced by the attractive forces of the planets. Now, if the plane of the orbit of the perturbed body coincides with that of the perturbing planet and if revolutions of the two are in the same sense, then the perturbing action will be strongest. If, however, the motion of the meteor family is retrograde and if the plane of the orbit is strongly inclined to that of the ecliptic, the perturbation will be much less. This in fact agrees with observation. Halley's comet has a retrograde motion and is much less disturbed than comets which travel direct. The major showers Leonids, Perseids, and Lyrids are highly inclined to the ecliptic (163°, 115°, 80°) and also travel in retrograde orbits. As such, they are very little perturbed and have been known to exist for a long time. The perturbing effect on a swarm of particles spreads the swarm out unevenly into a flat ring round the sun.

Since it takes time for the meteoric particles to be spread over the entire orbit by perturbing action, only showers of great age have such distribution. As such, showers which occur regularly annually may be presumed to be very

old. On the other hand, the members of a meteoric family which is of more recent origin are more or less located at certain places in the orbit.

The origin of the sporadic or non-shower meteor is still a matter of controversy—whether they belong to a solar system (elliptic orbit) or are extra-solar (hyperbolic orbit). The prevalent opinion is in favour of the former, at least, for a large majority of the cases. The point has to be determined from accurate determination of the velocities. When the meteor is at the sun-earth distance, its velocity in its orbit (heliocentric velocity), before being influenced by the gravitation attraction of the earth, must be less than 42·12 km./sec. (critical velocity) if it is to belong to the sun. Unfortunately, the methods of measurements, particularly the earlier ones, are not sufficiently accurate to decide the one or other. One significant fact in favour of elliptic orbits is that when a comparison of the heights of appearance and disappearance of shower meteors and non-shower or sporadic meteors is made no considerable difference in the heights is observed. [In a private communication (1950) to the author, Dr. A. C. B. Lovell of Manchester, England informs that he has an overwhelming mass of evidence (from observations only on meteors brighter than six magnitude) which shows that the percentage of meteors moving with hyperbolic velocities, even if any exist, is extremely small.]

An interesting and important point regarding the velocity of approach of the meteors and the variation of the frequency of meteoric impacts with the hour of the day and the geographical location of the place of observation may be noted here.

The velocity of approach of a meteor is its velocity relative to the earth. Since the earth is travelling in its orbit round the sun with a velocity 29 km./sec. with reference to the solar system, a meteor travelling with a velocity v with reference to the same system may appear to have any velocity within the range $v \pm 29$ km. per second. The maximum value $(v+29)$ will occur when the earth and the meteor happen to move towards each other and the minimum $(v-29)$ when they move away from each other. These limiting values are naturally of very rare occurrence. Generally, they move with a relative velocity given by $v \cos \psi_m + 29 \cos \psi_e$, where ψ_m and ψ_e are the angles which the velocity vector of the meteor and of the earth respectively make with the line joining the two. As the earth travels in space round the sun, one-half of its surface is always presented towards the direction of travel while the other half away from it. Evidently the forward half will encounter greater meteoric velocities than the rear half and hence have more frequent meteoric impacts.

Now, since the point on the forward hemisphere at which the zenith points to the direction of travel of the earth (in its orbit) will experience head-on collisions, the region round this point will be the one of greatest frequency of meteoric impacts. If the axis of the earth were at right angles to the plane of the ecliptic then in all seasons of the year this point would be on the equator where it is nearly 6-hour dawn. Since, however, the axis of the earth is tilted at an angle of $23\frac{1}{2}°$ to the plane of the ecliptic, the point

will wander between latitudes $\pm 23\frac{1}{2}°$ but will invariably be where it is nearly 6 hours before noon. In latitudes beyond the tropics the zenith at any point can never point to the direction of travel but will come nearest to this direction at nearly 6-hour dawn. It can thus generally be said that at any place the largest number of collisions with the maximum velocity may be expected to occur only when it is about 6-hour dawn at the place.

The above remarks apply only to sporadic meteors which arrive from random directions. In the case of shower meteors which appear to arrive from fixed radiants the highest number of impacts and the maximum velocities will be observed if and when the radiant happens to point to the direction of travel of the earth or nearest to it.

The composition of the meteors varies considerably from type to type. The chief constituents, as deduceed from meteor spectra are calcium, nickel, magnesium, manganese, chromium and, probably also sodium, aluminium and silicon. (See Fig. 2A, Plate I.) Meteor spectra may be classified into two distinct types: (i) Type Y showing strong H and K lines of ionized calcium, and (ii) Type Z where these lines are absent. Most of the shower meteors, specially the Leonids, appear to belong to class Y. Nearly two-thirds of sporadic meteors belong to class Z.

The meteorites falling on the earth are classified into two contrasting types—the iron meteorites and the stone meteorites. While the former contains very little of calcium, the latter contains more than one per cent of the same. Shower meteors belonging mostly to type Y (containing calciu m) may thus be classified as stone meteors. Non-shower or sporadic meteors, belonging to both type Y and type Z are thus partly stone meteors and partly iron meteors.

3. METEOR OBSERVATIONS—VISUAL AND PHOTOGRAPHIC

For applying the theory of meteors to the determination of density and temperature distribution in the middle atmosphere it is essential that the paths of the meteors be known accurately. The meteor paths may be measured if the same meteor can be seen (or photographed) by two observers suitably situated. Such observations yield the following data: Position of the radiant, the linear length of the path, the heights at the beginning and at the end of the luminous trail. For two observers the observational errors in all the above, except that for the radiant are also obtained. With more than two observers, the errors in the radiant, as also the personal errors of the observers, are known. In the best modern visual observations the errors are of the order 1° in radiant measurement and 2–4 km. in the height measurement.

Of the systematic meteor observations carried out in recent years mention may be made of those by the English workers [4], by the Harvard group in the U.S.A., by the American Meteor Society under C. P. Olivier, by individual co-operative bodies in Soviet Russia, by Hoffmeister in different latitudes and by McIntosh in New Zealand [5]. We give below some

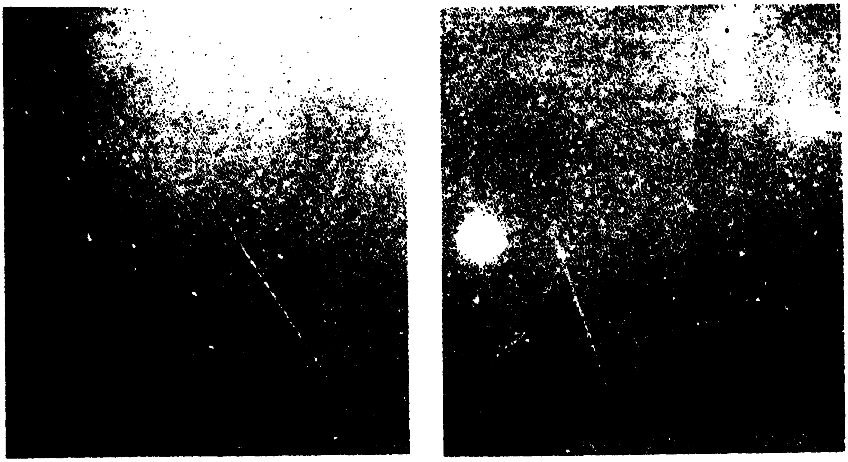


FIG. 2. Pair of photographs of the trail of Geminid meteor caught by cameras at Harvard and at Cambridge (U.S.A.) on Dec. 13, 1947. The breaks on the trails of the left picture are caused by rotating shutters.

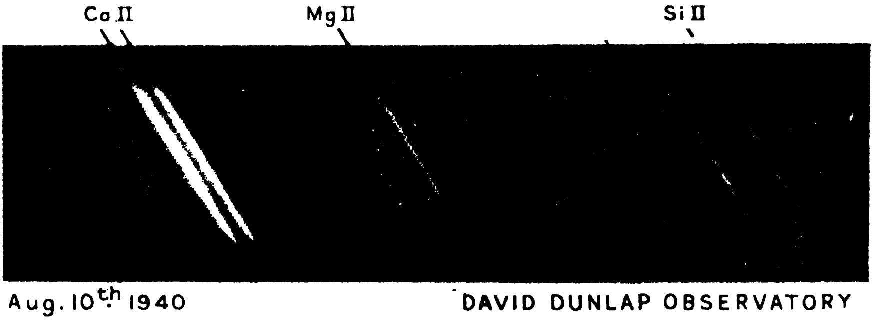


FIG. 2A. Spectrum of meteor trail showing the ionized elements Ca, Mg, Si. (After P. M. Millman.)

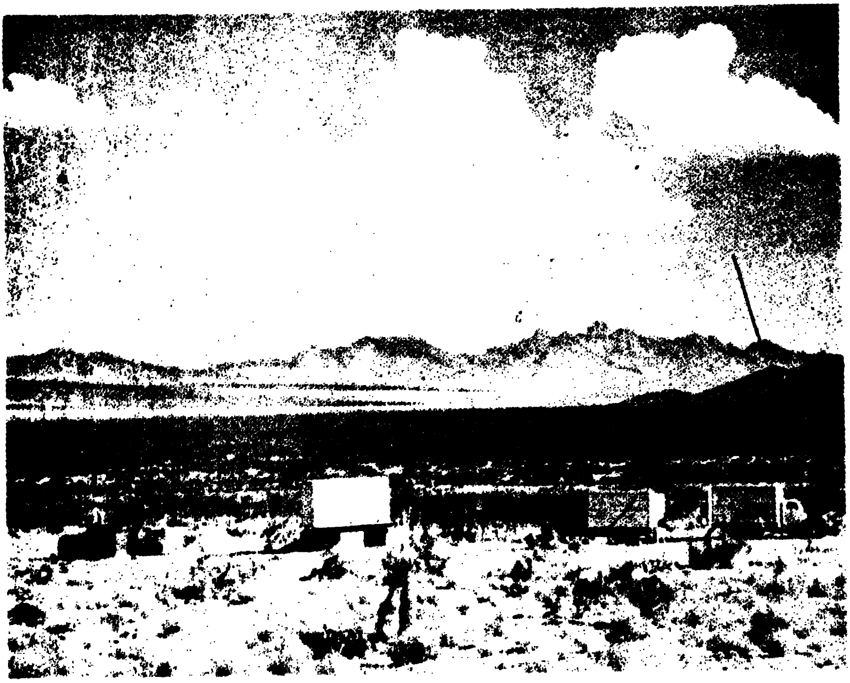


FIG. 3. One of the pair of mobile stations, Doña Ana, in New Mexico, U.S.A., for photographic study of meteors. Arrow indicates position of the other station (Soledad, hidden). (After F. L. Whipple.)

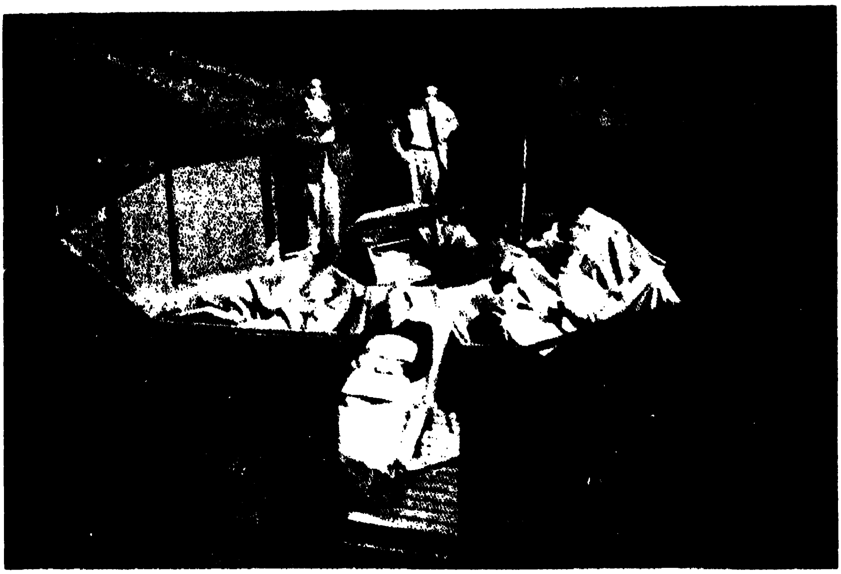


FIG. 3A. Method of making visual observations of meteors. National Research Council of Canada, Radio Field Station six miles south of Ottawa. Dominion Observatory photograph. (*Sky and Telescope.*)

details of the work by the Harvard group. (General accounts of the methods of making meteor observations are to be found in references [6].)

The first observations—visual—were made during the years 1931–33 by the Harvard-Cornell Meteor Expedition at Flagstaff, Arizona. The observers were E. Öpik, R. Wilson and D. Hargrave. 1436 meteor velocities were measured and subjected to statistical analysis. Similar observations were carried out by E. Öpik in 1934–38 at Tartu, Estonia [7]. But the most systematic observations are the photographic ones started under the auspices of the Harvard Photographic Meteoric Programme by F. L. Whipple in 1936 [8]. Up to the middle of 1948 the observating stations were located near Harvard, one at Cambridge and the other at Oak Ridge Station of the Harvard University with a base line of approximately 23 miles. The observing cameras were of 1·5-inch aperture and focal length 6 inches. Synchronous motors operated rotating shutters on the cameras to interrupt the exposure at intervals of 0·05 sec. The meteor trail, which is usually observable for less than a second, thus appeared on the photograph as a broken line. The position of any point on the trajectory can be measured with great accuracy when the meteor appears simultaneously on both the cameras. The separation of the successive breaks provide data for determining both the linear velocity and the deceleration due to the resistance of the atmosphere. The deceleration is generally small, of the order 1 km./sec.^{2} and was usually measurable only for the mid-point of the trail. In Fig. 2, Plate I, the pair of photographs shows the trail of Geminid meteor caught by the Cambridge and the Harvard cameras on December 13, 1947. The breaks on the trail in the photograph on the left due to the action of the rotating shutters are clearly visible.

By comparison of the photographed trail with the images of the stars, a rough measurement of the luminosity can be made. The photographs thus provide the following data of the meteors: heights of four points, luminosity (except at the end), the integrated luminosity, the velocity (this does not change greatly over the length of the trail), the deceleration at the middle point and the inclination to the direction of the zenith. Meteor data collected by the Harvard group have been utilized by Whipple to compute upper atmospheric data [9] (*vide infra*).

The meteor observing cameras were removed in 1948 to two chosen sites in New Mexico and are being operated there since then. The stations are separated by a distance of 12 miles and are situated at altitude of about 5,000 ft. above sea-level. Fig. 3, Plate II, shows one of the pair of the mobile stations. A super-Schmidt camera with aperture 12½ inches, focal length 8 inches, effective focal ratio 0·85 and field of angular diameter 52 degrees is under construction (1948). It is expected that with this camera the meteor 'yield' will be increased 40 times and slow meteors of 4th apparent visual magnitude will be photographed [10].

The heights of appearance and disappearance of meteors and the relations of these heights with the mass and the velocity of the meteor are of considerable importance in the theory of meteors. The following data

collected specially by Lindemann and Dobson in connection with their theory of meteors (*vide infra*) are therefore given.

The heights of appearance and disappearance of meteors lie within a wide range. The highest altitude at which meteors have been observed appear to be within Region *F* of the ionosphere. There is no lower limit of the height of disappearance since the falling of glowing meteors on the ground

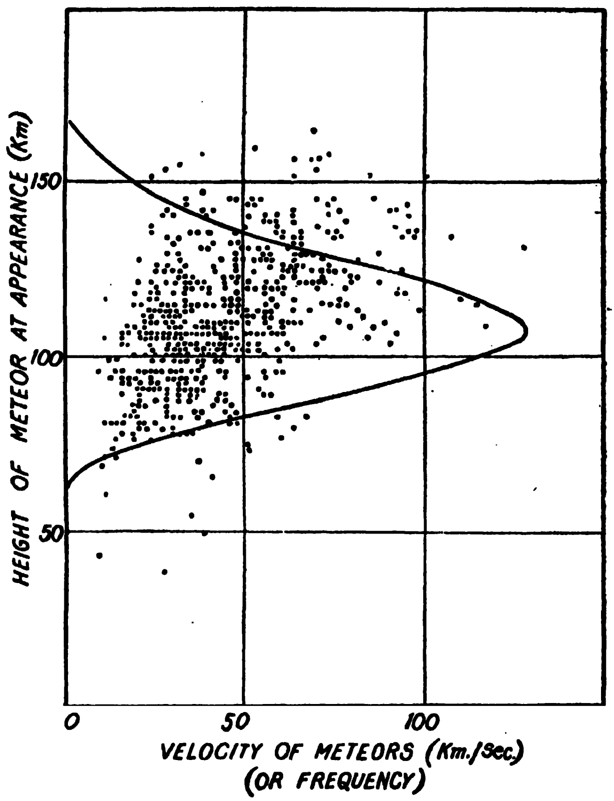


FIG. 4. Each dot represents one meteor, the ordinate giving its height at appearance and the abscissa its velocity. The curve shows the relative numbers of meteors appearing at different heights. (After Lindemann and Dobson.)

is not an uncommon phenomenon. Figs. 4, 5 represent the observed relation between the heights of appearance and of disappearance with frequency of occurrence of meteors. It will be noticed that while the heights of appearance roughly group near a single level at about 110 km., those of disappearance group about two distinct levels, one at 70 km. and the other at 48 km. There is a decided drop in the disappearance frequency curve at 55 km., where comparatively few meteors vanish. This latter fact is of considerable importance in the theory of meteors and will be referred to later.

A general relation exists between the heights of appearance and disappearance and the velocity of a meteor. A meteor having greater velocity appears higher up and also disappears higher in the atmosphere than one moving with a lower velocity [11].

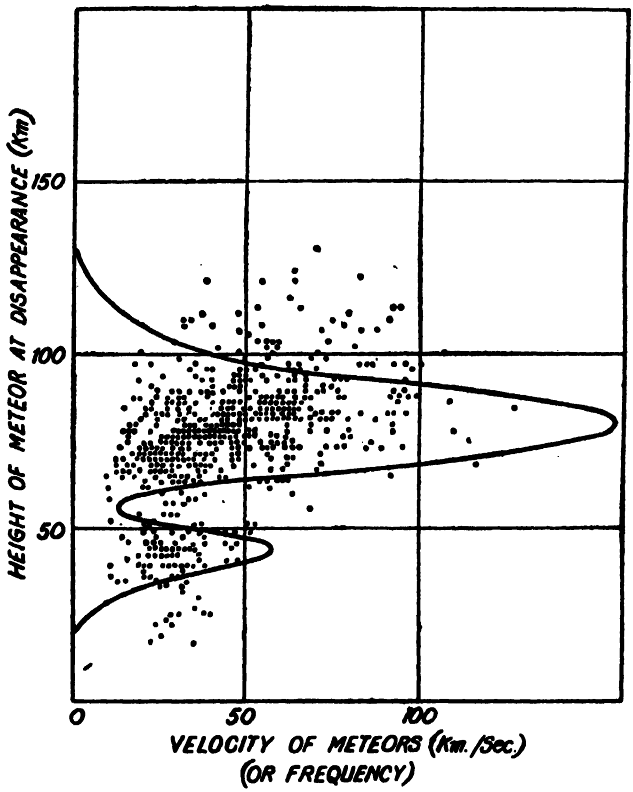


FIG. 5. Each dot represents one meteor, the ordinate giving its height at disappearance and the abscissa its velocity. The curve shows the relative numbers of meteors disappearing at different heights. (After Lindemann and Dobson.)

During the periods of principal showers of the year the mean heights of appearance and disappearance are found to increase. The possible reason of the greater heights of the showers is that shower meteors being stony volatilize at higher altitudes.

Excluding the showers and the abnormal ones which have low end points, a marked variation of heights with season is found [12]. It will be seen in Fig. 6 that the minimum height occurs in February and the maximum in August-September. From the figure it is also to be noticed that the range of meteors varies with season.

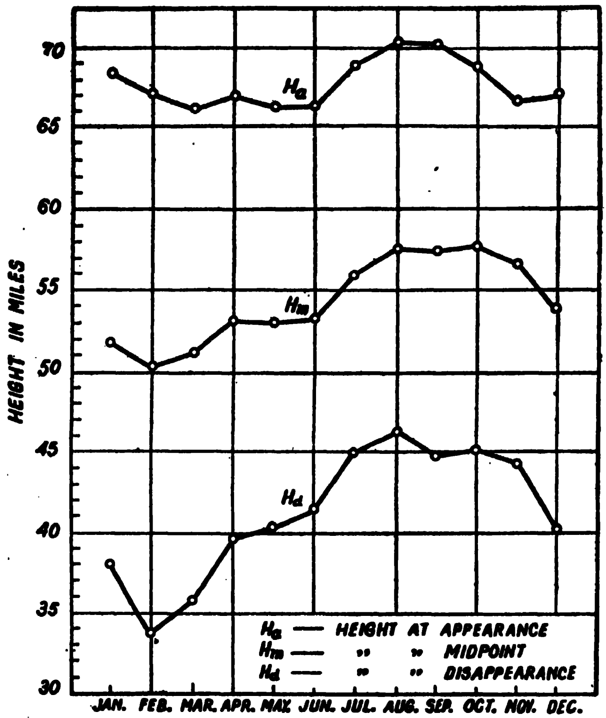


FIG. 6. Illustrating the seasonal variation of meteoric heights.

4. METEOR THEORIES

Many theories of the meteor have been proposed to explain the meteoric flashes, their heights of appearance and disappearance, distribution of light intensity along the trails, and other characteristics. Of these, the more important ones are those of Lindemann and Dobson (1923), of Sparrow (1926) and of Öpik (1933).

As already mentioned Lindemann and Dobson were the first to develop a complete theory of the meteor and to predict therefrom the existence of a region of high temperature in the middle atmosphere. But, notwithstanding the success of the prediction, the theory has been subject to criticisms. These will be referred to later.

In the present section we will first discuss the Lindemann-Dobson theory in some detail, because, though criticized, the Lindemann-Dobson theory is of lasting value being the first attempt of its kind to clarify the process of encounter of a meteor particle with air molecules. Accounts of Sparrow's theory and of Öpik's theory (the latter after a treatment of Herlofson), will then follow. Finally, F. L. Whipple's analysis of meteor data and results of his calculations on upper atmospheric temperature and density will be given.

(a) Lindemann-Dobson theory

The fundamental assumption of Lindemann and Dobson is that as a meteoric particle rushes into the earth's atmosphere, a hot gas cap is formed immediately in front of it due to the adiabatic compression of the air. Heat from the gas cap is transmitted to the meteor and when the radius of the meteor is small, as is the case with those which disappear at great heights, the entire mass is raised to a high temperature and the meteor becomes incandescent. A knowledge of the meteoric data such as the velocity, density, radius, heat-conductivity and specific heat, enables one to calculate the atmospheric density at the height of appearance or of disappearance of the meteor. The calculations for the height of appearance are independent of those for the height of disappearance and both the calculations show that the estimated densities cannot be reconciled with a low temperature in the region of average meteoric display—but are consistent with a high temperature of the order of 300°K.

It is to be noted that the calculations of Lindemann and Dobson refer to ordinary meteors. Very bright meteors like fire-balls which come close to the surface of the earth and are much larger than the average size are excluded. The average meteor of Lindemann and Dobson has the following characteristics: height of appearance—100 km.; height of disappearance—80 km.; length of the path traversed (obliquely)—60 km.; speed—40 km. per second. It is thus visible for a period, t_0, of about 1·5 sec. The apparent brightness at a distance of 150 km. is equal to that of a first magnitude star, i.e., energy radiated per second is $E = 3{\cdot}3 \times 10^{10}$ ergs. This latter quantity also determines the size of the average meteor. The total energy dissipated in the period 1·5 sec., during which the meteor is visible, is $1{\cdot}5 \times 3{\cdot}3 \times 10^{10} = 5 \times 10^{10}$ ergs. This energy is derived from the kinetic energy of the meteor $\frac{1}{2}mv^2$ (average $v = 4 \times 10^6$ cm./sec.). Equating the two energies, m, the mass of the meteor, is found to be $6{\cdot}25 \times 10^{-3}$ gm. If the meteor is composed of iron, its radius (assuming spherical shape) is 0·057 cm.

(i) *Density required for the formation of the gas cap.*—During its flight through the atmosphere a meteor at first experiences direct impacts by the individual air molecules, the mean free paths of which at great heights are much larger than the dimensions of the meteor itself. As the meteor descends it enters into regions of gradually increasing density and, at a height depending on the size of the meteor and its velocity, a protecting layer of compressed air begins to form in front of it. Direct impact on the surface of the meteor by the air molecules now ceases. The atmospheric density at which this will happen may be estimted as follows: Suppose that, to start with, the meteor, by its forward velocity, has trapped a molecule against its surface. This trapped molecule, owing to its mean thermal velocity, will in general move out of the region swept by the surface of the meteor within a certain time unless this time be equal to or greater than the time required by the meteor to cover the distance of a mean free

path of the molecules. In the latter case other molecules will collect in front of the meteor before the trapped molecule finds time to escape and a protecting layer begins to form. We thus have for the condition of formation of the layer,

$$L_1 = r \cdot \frac{v}{V_0},$$

where L_1—mean free path of the molecule,
r—radius of the meteor,
V_0—mean velocity of the molecule,
and v—velocity of the meteor.

If ρ_1 be the density at which the above condition holds and L_n and ρ_n be the mean free path and density at normal temperature and pressure, then we have

$$\rho_1 = \frac{L_n}{r} \cdot \frac{V_0}{v} \cdot \rho_n. \qquad \ldots \qquad \ldots \qquad \ldots \qquad (1)$$

Hence the gas cap begins to form when the density of the atmosphere is of the order of $L_n V_0/rv$ times the normal density. Thus, in the case of nitrogen, direct impacts cease and the gas cap begins to form at the level at which the density is given by $5{\cdot}6\times10^{-4}/rv$. The value of v may be obtained from direct observations on the meteor and that of r from the formulae $r = (3m/4\pi\rho_m)^{1/3}$ and $\frac{1}{2}mv^2 = Et_0$, as shown above, where ρ_m is the density of the meteoric material.

The quantity of heat available for raising the temperature of the meteor by direct molecular impacts is assumed to be negligible. It is only when the gas cap forms that the heat from this cap is partly transmitted to the meteor.

(ii) *Rate of loss of energy by the meteor.*—During its flight the meteor drives the air molecules in its path with its own velocity and also compresses adiabatically the air in front of it from the initial pressure p_0 to a pressure p, say. The total kinetic energy abstracted from the meteor per second is therefore utilized partly in imparting kinetic energy to the air molecules in front of it and partly in supplying the potential energy of the adiabatically compressed air cap.

The kinetic energy gained per second by the trapped air molecules is equal to

$$\tfrac{1}{2}m'v^2 = \tfrac{1}{2}(S\rho_0 v)v^2 = \tfrac{1}{2}S\rho_0 v^3,$$

where m'—mass of the trapped air molecules,
S—cross-sectional area of the meteor,
ρ_0—atmospheric density at the height concerned.

The potential energy gained per second by the air molecules is the amount of work done in compressing a mass $S\rho_0 v$ of the gas adiabatically from the pressure p_0 to the pressure p. This is equal to

$$m'\left(\frac{RT_0}{M_0}\right)\log p/p_0$$

$$\frac{S\rho_0 vV^2}{\gamma}\cdot\log p/p_0,$$

where R—universal gas constant,

T_0—temperature of the atmosphere,

M_0—molecular weight,

V—velocity of sound in air,

γ—ratio of specific heats of air,

p—pressure of the gas cap.

The total energy lost by the meteor per second in imparting the kinetic and potential energies is therefore given by

$$E' = \tfrac{1}{2}S\rho_0 v^3 + \frac{S\rho_0 vV^2}{\gamma}.\log p/p_0$$

$$= \tfrac{1}{2}S\rho_0 v^3\left[1+2/\gamma.\left(\frac{V^2}{v^2}\right)\log p/p_0\right]$$

$$= \tfrac{1}{2}S\rho_0 v^3, \qquad \qquad (2)$$

the second term within square brackets being small compared to unity.

To determine the pressure p of the gas cap we note that the resistance experienced by the particle is $p.S$ and the work done by it against this resistance per second is $p.S.v$. Thus

$$E' = p.S.v = \tfrac{1}{2}S\rho_0 v^3$$

$$\text{or} \quad p = \tfrac{1}{2}\rho_0 v^2. \qquad \qquad (3)$$

(iii) *Energy available for heating the meteor: Efficiency factor k.*—The total work done by the meteor per second in moving a distance v is $\tfrac{1}{2}S\rho_0 v^3$. Of this, only a small fraction is available for heating the meteor because the molecules of the gas cap which drift behind its side carry with them a large fraction of the energy. In fact, since a certain proportion of the constituent molecules of the gas cap is always escaping from it we may suppose that the gas cap itself is not moving exactly with the velocity v but with a smaller velocity.

If now V' be the relative velocity of the meteoric particle with respect to the gas cap ($V'<v$), the amount available for heating will be $\tfrac{1}{2}S\rho_0 v^2.V'$. The value of V' may be determined from the fundamental concepts of the kinetic theory of gases and may be shown to be equal to $(V_1—V_2)/3$, where V_1 is the velocity of the compressed gas molecules in front of the meteor and V_2 that of the gas molecules at the temperature of the meteoric surface.

Hence the heat which flows to the meteor per second is equal to

$$\tfrac{1}{2}S\rho_0 v^2 . V' = \tfrac{1}{2}S\rho_0 v^2 . \frac{V_1-V_2}{3}.$$

If we denote by k the fraction of the total work done which is utilized in heating the meteor, we get

$$k = \frac{\frac{1}{2}S\rho_0 v^2(V_1 - V_2)/3}{\frac{1}{2}S\rho_0 v^3} = \frac{V_1 - V_2}{3v} \quad \cdots \quad \cdots \quad (4)$$

Lindemann and Dobson call this fraction k the efficiency factor of heating. Expressed in terms of temperatures,

$$\begin{aligned} k &= \frac{1}{3}\left(\frac{V_1}{v} - \frac{V_2}{v}\right) \\ &= \frac{1}{3}\left(\sqrt{\frac{T_1}{T_v}} - \sqrt{\frac{T_2}{T_v}}\right) \quad \cdots \quad \cdots \quad \cdots \quad (5) \end{aligned}$$

where T_v is a fictitious temperature defined by $T_v = M_0 v^2/3R$, T_1 is the temperature of the gas cap and T_2 the temperature of the surface of the meteor. T_1 is given by

$$\frac{T_1}{T_0} = (p/p_0)^{\frac{\gamma-1}{\gamma}}$$

$$\text{or} \quad T_1 = T_0\left(\frac{\rho_0 v^2/2}{\rho_0 V_0^2/3}\right)^{\frac{\gamma-1}{\gamma}} = T_0\left(\frac{3v^2}{2V_0^2}\right)^{\frac{\gamma-1}{\gamma}}, \cdots \quad \cdots \quad (6)$$

where T_0 is the initial temperature of the gas.

(iv) *Density at the height of appearance.*—The fraction k of the total amount of work done per second utilized in heating the meteor as obtained above, may be used for finding out the density at the level of the atmosphere at which a meteor will make its appearance. Heat is supplied to the meteor from the gas cap at the rate of $k \cdot \frac{1}{2}\rho_0 v^3$ per unit surface per second. This is carried into the interior of the meteor at the rate of $K. dT/dr$, where K is the thermal conductivity of the meteoric material and dT/dr the temperature gradient.

Now evaporation will take place when the heat supplied is of the order of (or greater than) the heat conducted into the meteor, i.e., when

$$k \cdot \tfrac{1}{2}\rho_0 v^3 = K\frac{dT}{dr},$$

$$\text{or} \quad \frac{dT}{dr} = \frac{k \cdot \frac{1}{2}\rho_0 v^3}{K}. \quad \cdots \quad \cdots \quad \cdots \quad (7)$$

It can be shown from the above equation that evaporation will begin in case of meteors of the average small size (radius $\fallingdotseq 6\times10^{-2}$ cm. as deduced above), when the temperature at the centre is almost equal to that at the surface. Under this condition

$$\int ms\, dT_2 = \int \pi r^2 k \cdot \tfrac{1}{2}\rho_0 v^3\, dt, \cdots \quad \cdots \quad \cdots \quad (8)$$

where m—mass of meteor,

s—specific heat of its material.

The integral on the right hand side is the rate at which heat flows into the meteor of radius r.

For a meteor descending vertically, $v\,dt = -dh$, where h is the height measured from the ground level upwards.

Thus

$$msT_2 = -\tfrac{1}{2}\pi r^2 v^2 \int k\rho_0\, dh.$$

For convenience we henceforth denote the density at the ground level by ρ_0 and that at any height by ρ. Now, since $m = \frac{4}{3}\pi r^3 \rho_m$, ρ_m being the density of the meteoric material, we have

$$\frac{8}{3}\frac{\rho_m s T_2 r}{v^2} = -\int k\rho\, dh.$$

The mean value of k between the initial and final values of the surface temperature of the meteor is obviously $k/2$, where k is the efficiency factor at the final temperature T_2. Taking this mean value as constant,

$$\frac{16}{3}\frac{\rho_m s T_2 r}{kv^2} = -\int \rho\, dh.$$

Let ρ_a be the density at the height h at which the temperature of the whole meteoric mass attains the value T_2, i.e., at which the meteor evaporates copiously and consequently becomes visible. Then since

$$\rho_a = \rho_0 \exp\left(-\frac{gM_0 h}{RT_0}\right)$$

and since the atmospheric density at which the cap begins to form is negligible compared to that at the height of appearance, we have

$$-\int \rho\, dh = \rho_a \frac{RT_0}{gM_0}. \qquad \text{(9)}$$

Hence

$$\rho_a = \frac{16}{3}\frac{\rho_m s T_2 r}{kv^2}\frac{gM_0}{RT_0}. \qquad \text{(10)}$$

For meteors descending at an angle χ to the vertical

$$\rho_a = \frac{16}{3}\cdot\frac{\rho_m s T_2 r \cos\chi}{kv^2}\cdot\frac{gM_0}{RT_0}. \qquad \text{(11)}$$

All the quantities on the right hand side of this equation can be ascertained fairly accurately with the exception of T_0, the temperature of the upper atmosphere. ρ_a can thus be calculated if a value of T_0 is assumed.

(v) *Density at the height of disappearance.*—As already mentioned the density can also be calculated from considerations of quite a different nature about the height of disappearance of meteors.

Once the surface of the meteor is raised to such a temperature that evaporation takes place, practically all the heat reaching the meteor is used up in volatilizing it. Thus

$$l\,dm = \pi r^2 . k . \tfrac{1}{2}\rho v^3\, dt \qquad \text{(12)}$$

where l is the latent heat of evaporation of the meteoric material and dm is the amount evaporated in time dt.

It can be shown that deceleration of meteors (excepting the very slow ones) is practically negligible, so that V_1, the mean velocity of the atmospheric molecules in the gas cap can be treated as constant. Again, since the temperature of the meteoric surface remains sensibly constant so long as the meteor glows, V_2, the mean velocity of the molecules receding from the meteor does not vary appreciably. Hence unlike the value of k in the case previously considered, k here is constant being given by

$$k = \frac{V_1 - V_2}{3v}.$$

Since $m = \frac{4}{3}\pi r^3 \rho_m$, we have

$$dm = 4\pi r^2 \rho_m dr.$$

Hence, dr/dt, the rate at which the size of the meteor decreases due to evaporation, is given by

$$\frac{dr}{dt} = \frac{1}{4\pi r^2 \rho_m} \cdot \frac{dm}{dt}$$

$$= \frac{1}{24} \frac{V_1 - V_2}{l} \cdot \frac{\rho v^2}{\rho_m} \quad \ldots \quad \ldots \quad \ldots \quad (12.1)$$

from Eq. (12).

If L is the total length of the meteor trail making an angle χ with the vertical

$$v\,dt = \frac{dh}{\cos\chi} = \frac{L}{\Delta h} \cdot dh,$$

where Δh is the projection of L on the vertical.

Hence, from Eq. (12.1),

$$dr = \frac{1}{24} (V_1 - V_2) \frac{v\rho L}{l\rho_m} \cdot \frac{dh}{\Delta h}.$$

Thus r, the radius of the meteor when it started to evaporate, is given by

$$r = \int_r^0 dr = \frac{1}{24} (V_1 - V_2) \frac{vL}{l\rho_m \Delta h} \int_{\rho_a}^{\rho_d} \rho\, dh$$

$$= \frac{V_1 - V_2}{24} \cdot \frac{vL}{l\rho_m \Delta h} \cdot \frac{RT_0}{gM_0} (\rho_d - \rho_a), \quad \ldots \quad \ldots \quad (13)$$

where ρ_d is the atmospheric density at which the whole of the meteor mass evaporates. If ρ_a is neglected in comparison with ρ_d it follows that

$$\rho_d = \frac{24r}{V_1 - V_2} \cdot \frac{l\Delta h}{vL} \cdot \rho_m \cdot \frac{gM_0}{RT_0} \quad \ldots \quad \ldots \quad \ldots \quad (14)$$

The atmospheric density at the height at which a meteor of radius r disappears can be estimated from this equation.

It is evident that these equations will hold good only for an isothermal atmosphere. That is, the meteor, during its whole course, is assumed to pass through an atmosphere at constant temperature.

(vi) *Results of density distribution calculation: Indication of high temperature in the middle atmosphere.*—Lindemann and Dobson utilized the data obtained from particulars of several hundreds of meteoric observations collected by W. F. Denning and his co-workers (Figs. 4 and 5) for calculating, with the help of equations (11) and (14), the densities at the heights of appearance and disappearance respectively of the meteors. The results

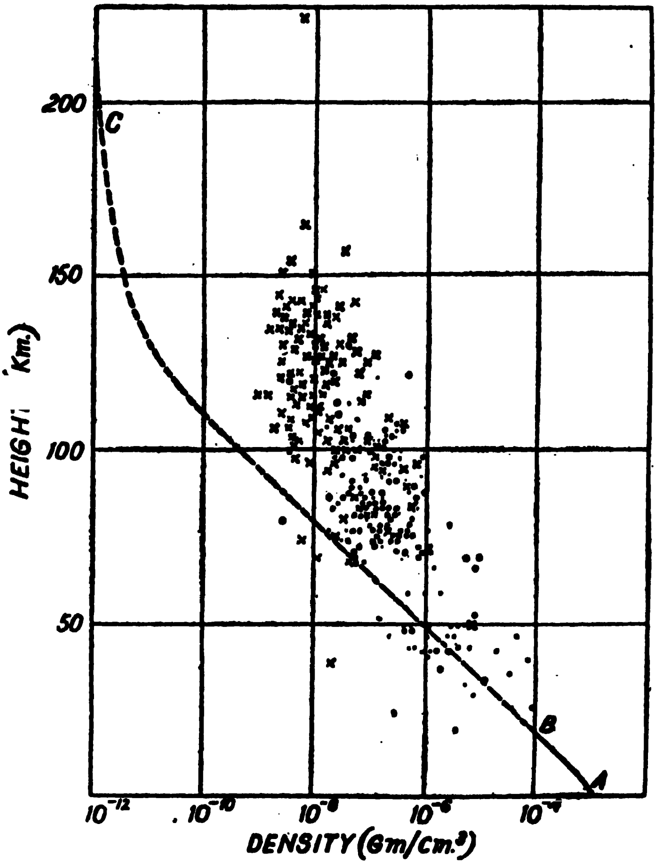


FIG. 7. Plot of density against height obtained from meteoric data. The crosses represent the density and height at the points of appearance and the dots those at the points of disappearance of the meteors.

The broken line curve depicts the variation of density with height on the assumption of an isothermal atmosphere at 220°K and preponderance of lighter gases above 100 km.; the continuous portion of the line at the bottom (*AB*) is obtained from sounding balloon observations. (After Lindemann and Dobson.)

represented by Fig. 7 shows that for the region 60 km. to 160 km. (within which a large amount of observational data was available) the calculated densities are much higher than those obtained directly from the formula

$$\rho_h = \rho_0 \exp\left(-\frac{gMh}{RT_0}\right)$$

on the assumption of an isothermal atmosphere at a constant low temperature—220°K. The results are, however, consistent with a higher temperature—say, 300°K—the difference in densities in fact amounting to 1,000 times. Further, the values obtained from the 'appearance' considerations and those obtained from the 'disappearance' considerations support each other. It is to be noted that for calculation of the densities with the help of Eqs. (11) and (14) a knowledge of the temperature itself is necessary. This can, however, be found out from the slope of the height-density curve (densities calculated with any assumed temperature) drawn from a number of observations. It is, of course, assumed that the region under consideration is an isothermal one.

The above conclusion regarding the existence of a high temperature in the middle atmosphere receives support from three other associated meteoric phenomena discussed below.

The dip at 50–60 km. in the height-disappearance frequency curve (Fig. 5) is explained by Lindemann and Dobson as due to the presence of an isothermal region of high temperature above 60 km. If the atmosphere through which the meteors are rushing be isothermal down to the heights of disappearance, then the dots representing the latter ought to group round a single maximum as in the case of the height-appearance frequency curve (Fig. 4). If, however, there be a region of cooler atmosphere which the meteor has to pass through in its downward course, then its rate of evaporation will be checked and the meteor would not disappear at the height at which it would have done normally. It will come down to a level the density at which is high enough to reinstate its rate of heating and evaporation. The presence of a minimum in the curve in Fig. 5 thus indicates the existence of a lower temperature below 55 km. than above. (It has, however, been pointed out by F. L. Whipple that the dip is an evidence of high temperature rather *in* the 50–60 km. region than above it. 'The existence of a smaller logarithmic density gradient prolongs the lifetime of a meteor reaching this height; the meteor has, therefore, a smaller chance of disappearing in the critical range of altitude. The air temperature, *per se*, is not the determining factor, but rather the consequent reduced density gradient.')

The height at which the meteor is first seen is certainly less than the height at which the gas cap, to which the heating and the incandescence of the meteor are due, first begins to be formed. The density at the latter height, which can be calculated with the help of the simple relation given by Eq. (1), might therefore be considered as a lower limit of the probable density at the actual height of appearance of the meteor. Even when these lower estimates of the desity and the corresponding heights of appearance are plotted, the height-density curve so obtained is found (Fig. 8) to be inconsistent with an isothermal atmosphere at 220°K but consistent with an atmosphere at a much higher temperature. (See, however, the criticisms below.)

It can be shown that there is a minimum velocity below which no meteoric flash should occur. This is because a lower velocity would not

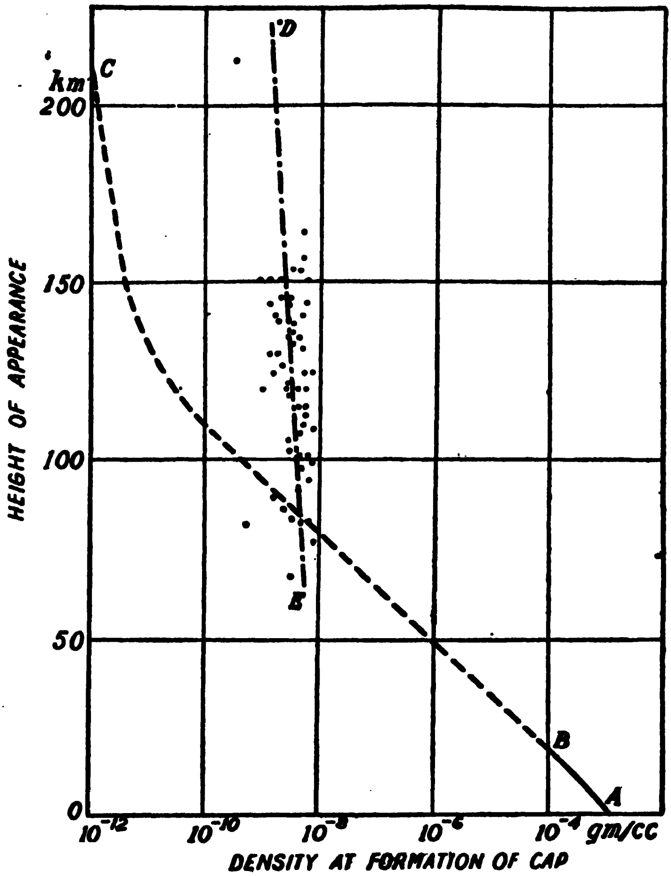


FIG. 8. Each dot represents the lower limit of the density required for the formation of gas cap in front of a meteor at the height of its appearance.

Curve *ABC* is the same as in Fig. 7. The line *DE* drawn through the dots shows that the density distribution corresponds to an isothermal atmosphere at a temperature much higher than 220°K. (After Lindemann and Dobson.)

cause appreciable heating and hence the evaporation would be too small to make the meteor visible. Calculations show that for a temperature of 220°K in the region of appearance of meteors this minimum velocity is 19 km./sec., while for a temperature of 300°K the minimum is 12 km./sec. Observations show, however, that a large number of meteors have velocities below 19 km./sec., thus indicating a temperature higher than 220°K.

The existence of high temperature in the middle atmosphere as deduced from the study of abnormal propagation of sound is thus confirmed by the critical study of meteoric flashes.

We shall see in the next chapter that the rise of temperature in the middle atmosphere is due to the presence of ozone. Ozone strongly absorbs

the near ultraviolet portion of the solar spectrum (2200Å to 2900Å) and thus forms, as it were, a heat reservoir in the middle atmosphere. The coefficient of absorption is so high that though the ozone is distributed between 15 and 45 km. most of the absorption and consequent heating occurs in the topmost portion of the ozone layer.

(vii) *Criticisms of Lindemann-Dobson Theory.*—As already mentioned the theory has been subject to criticisms on various grounds. The first of these were from Sparrow [13] and from Epstein [14] on thermodynamical grounds. But these were met by Lindemann [15]. Later, Herlofson [16] has raised objections regarding the mode of heating by which the meteor is raised to incandescence. According to Herlofson, 'the assumption that the meteor will be heated by adiabatic compression only is not permissible and leads to a serious underestimate of the amount of heat created in the meteor and the cap'. In support of this it is pointed out that the conclusion of Lindemann and Dobson, as deduced from their theory, that the air density in the region 100–150 km. is not less than 10^{-9} gm./cm.^{3} (see Fig. 8) is found to be wrong by a factor of 100, from ionospheric measurements in this region (see Chapter VI).

Further, 'a closer analysis of the collisions between the resting air molecule in front of the meteor, and air molecules which have hit the meteor surface once, indicates that an air cap may be formed when the mean free path is comparable with the dimensions of the meteor, and not with the dimensions multiplied by the ratio between the velocity of the meteor and the thermal velocity of the air molecules'. Hence it follows that 'an adiabatic air cap cannot be expected to form in front of meteors even as large as some centimeters, unless they penetrate to heights well below 100 km., and for ordinary shooting stars, whose dimensions are probably 0·01–1 cm., an air cap can hardly be supposed to form above heights of the order 50–70 km.' If, however, a meteor is able to reach levels where the mean free path is substantially smaller than the meteor dimensions, the equations employed by Lindemann and Dobson become valid. It is for this reason that estimate from Lindemann-Dobson theory approaches correct values at heights lower than 50 km. (For a fuller discussion Herlofson's original paper [16] should be consulted.) From the remarks made above it is clear that for the higher regions of the atmosphere, where by far the largest number of meteors appear and disappear, one has to consider the primary effect, namely the effect of direct impacts of single air molecules with the body of the meteor. This was first done by Sparrow and is discussed below.

(*b*) Sparrow's theory

The essential features of Sparrow's theory are quite simple. The meteor in rushing forward experiences direct impacts of the air molecules upon its surface in rapid succession. Each of these impacts imparts some energy to the meteor which appears as heat. Equating the rate of absorption of energy due to these impacts to the rate of radiation according to Stefan's

law, the temperature of the meteoric particle is obtained. When this temperature is high enough to produce copious evaporation, the meteor becomes visible.

(i) *Efficiency factor k.*—The collision of an air molecule with the meteor is really between an air molecule and a meteoric molecule. Sparrow considers the collision to be inelastic since there are several factors which will cause dissipation of the energy. To find out the fraction k of the total energy of impact received by the meteor we consider the meteoric molecule to be at rest and the air molecule to strike it with a velocity v relative to it.

Consider first the simple case in which the direction of approach is along the line joining the centres of the two meteors. It is supposed that as a result of collision the relative velocity (v) is reduced to zero. If m_1 and m_2 be the masses of the air and the meteor molecules respectively, then, in order that the momentum may be conserved, the system, air-meteor molecule will have, after collision, velocity equal to $m_1v/(m_1+m_2)$. The energy of the system, $\frac{1}{2}(m_1+m_2)m_1v^2/(m_1+m_2)^2$ will be less than the original energy $\frac{1}{2}m_1v^2$ by the amount $\frac{1}{2}m_1v^2m_2/(m_1+m_2)$. This energy will have to be dissipated.

If we take the more general case, in which the direction of approach makes an angle θ with the line joining the centres of the molecules, then supposing that the collision reduces the relative velocity along the line of centres to zero while it leaves unaffected the velocity perpendicular to the line of centres it is easily shown that after impact the energy of the air molecule is

$$\frac{1}{2}m_1v^2\left\{1-\frac{2m_1m_2+m_2^2}{(m_1+m_2)^2}\cos^2\theta\right\},$$

and the energy of the meteoric molecule is

$$\frac{1}{2}m_1v^2.\frac{m_1m_2}{(m_1+m_2)^2}\cos^2\theta = q_1\cos^2\theta\text{, say.} \quad .. \quad .. \quad (15)$$

The sum of these two energies is less than the original energy by the amount

$$\frac{1}{2}m_1v^2\,\frac{m_2}{m_1+m_2}\cos^2\theta = q_2\cos^2\theta\text{, say.} \quad .. \quad .. \quad (16)$$

This energy is therefore dissipated.

The dissipated energy may be used up in each of these three ways: (*a*) dissociation of the atmospheric molecule, (*b*) dissociation of the meteoric molecule or disruption of its lattice structure, and (*c*) ionization of the constituent atoms. Of these only the portion of energy used up in method (*b*) is available for heating the meteor.

It is assumed that besides the gain of kinetic energy, the meteoric molecule also acquires half of the dissipated energy to its credit. Thus the net gain of energy is $(q_1+\frac{1}{2}q_2)\cos^2\theta$. The ratio of this latter to the total energy of impact $\frac{1}{2}m_1v^2$ is called by Sparrow the efficiency factor k. To take into account all possible directions of approach defined by θ, the

mean value of $\cos^2\theta$ for all collisions is required. Since this is equal to 1/2, the efficiency factor

$$k = \frac{1}{2} \cdot \frac{(q_1 + \frac{1}{2}q_2)}{\frac{1}{2}m_1 v^2};$$

or, from Eqs. (15) and (16),

$$= \frac{m_2^2 + 3m_1 m_2}{4(m_1 + m_2)^2}. \quad \ldots \quad \ldots \quad \ldots \quad (17)$$

Table II gives the values of k calculated for hydrogen and nitrogen molecules striking a meteor. Two different values of m_2, 25 and 56, are used.

TABLE II

Values of k

	Hydrogen	Nitrogen
$m_2 = 25$	0·27	0·37
$m_2 = 56$	0·26	0·27

(ii) *Relation between velocity and height of appearance of meteors.*—Since v is the velocity of the meteor, it is struck by $\pi r^2 nv$ molecules per second where n is the number of molecules per cm.^{3} and r the radius of the meteor. The total energy gained per second by the meteor is therefore

$$\pi r^2 nvk \,.\, \tfrac{1}{2}m_1 v^2 = \frac{\pi}{2}\, k\rho v^3 r^2,$$

where ρ is the density of the gas at the height concerned. Assuming the meteor to be a black body radiating at a temperature T°K the energy radiated per second is $4\pi r^2 \sigma T^4$, where σ is the Stefan's constant.

Equating the energy loss to the energy gained per second,

$$4\pi r^2 \sigma T^4 = \frac{\pi}{2}\, k\rho v^3 r^2,$$

$$\text{or} \qquad T^4 = kv^3\rho/8\sigma. \quad \ldots \quad \ldots \quad \ldots \quad (18)$$

In this equation k, v and σ are known and T may be assumed to be of the order of 2000°K, being the temperature of copious evaporation, i.e., the flashing point of the meteor. ρ can thus be calculated for the height of appearance of the meteor, which is obtained from direct observation. The density at this height being thus known, that at any other height can be at once computed assuming an isothermal atmosphere. Hence the distribution of density with height is fixed once for all. Equation (18) then predicts a definite height of appearance corresponding to a particular value of v. The velocity and the corresponding height of appearance calculated in this way seem to agree well with observations.

Sparrow's theory of meteors as sketched above is obviously not as completely developed as the theory of Lindemann and Dobson. Further,

the theory has been criticized by Lindemann on various grounds [15]. Nevertheless the fundamental idea of Sparrow, namely that the ultimate source of heating is the direct impact of air molecules with the meteor surface, has been the basis of all subsequent discussions regarding the majority of meteors which appear in the high atmosphere as shooting stars. This, in particular, refers to Öpik's work which we now proceed to describe.

(c) Öpik's theory—Herlofson's Treatment

Öpik's has developed a detailed theory of meteors. For a full account of the theory the reader is referred to Öpik's original papers [17]. Here, the main principles of the theory will be applied to a simplified model after Herlofson [16].

The essential problem in Sparrow's theory of meteors is the effect of collision of an air molecule at rest with the surface of a solid body (composed of iron for instance) moving with a velocity of 20 to 70 km./sec. The kinetic energy of air molecules (N_2 and O_2) for this relative velocity corresponds to 60 to 800 eV. (For oxygen and nitrogen atoms the amounts are halved.) Unfortunately, there are no experimental results available on the effect of impact with solid surfaces of air molecules with such energies. There are, however, experimental studies on impact of fast moving positive ions with metal surfaces, and, Öpik has developed his theory with the results of such experiments as guide.

Remembering that the binding energy of an atom in solids is only a few electron volts, two possible results of impact of air molecules with meteor surface suggest themselves. One is that the impinging air molecule ejects (that is, causes evaporation of) a meteor molecule. The phenomenon is analogous to that of 'sputtering' in discharge tubes. Experiments on sputtering show, however, that for the energy range under consideration, for each impinging molecule, only a fraction of an atom can be evaporated [18]. The 'sputtering effect' cannot therefore possibly be the dominant factor in the energy balance of a meteor. The other possibility is that the impinging molecule is trapped in the meteor surface. This effect is more promising because experiments on accommodation coefficient for positive ions of noble gases [19] show that a major proportion of the impinging ions (1/2 to 3/4) are trapped in the metal surface where they necessarily give up their kinetic energy. Öpik supposes that an analogous process occurs in the case of impact of air molecules (neutral) with the meteor surface.

It is assumed that in the region of meteoric flash, the mean free path of the molecule is much greater than the radius of the meteor. Under such condition, the front surface of the meteor is bombarded by single air molecules and the major proportion of them is trapped in the metal surface. The kinetic energy (relative to the meteor) which is given up to the meteor as heat suffices to bring the temperature to that of evaporation. The meteor atoms (mostly iron) thus distil off the meteor with velocity appropriate to the temperature.

The distinction between Sparrow's process and the process described above should be clearly understood. In Sparrow's process the energy of the impinging air molecules is spent partly in heating the meteor and partly in ejecting a meteoric atom and imparting to it a comparatively large kinetic energy with respect to the meteor. In the process outlined above the energy (with respect to the meteor) of the atoms which are evaporated is very small being only the thermal energy. The evaporated atoms have thus practically the same velocity as that of meteor and when they collide with air molecules produce light and ionization. It is to be noted that according to the above, the energy supplied by each air molecule on collision is enough for evaporation of a great number of atoms because the binding energy of the meteor atom in the solid body amounts only to few electron volts. For complete evaporation therefore, the total air mass intercepted is small compared to the mass of the meteor body. The meteor, in course of its passage through the atmosphere, is not appreciably retarded. The atmosphere therefore does not really check the flight of the meteor but rather breaks the single meteor into atoms and disperses them along the meteor track.

The equation expressing the balance between the heat supplied by the impinging air molecules and the latent heat carried away in the evaporation may be expressed thus on the assumption that the collision is inelastic and the molecule trapped on the surface gives up the whole of its kinetic energy as heat:

$$4\pi r^2 \rho_m l dr = \tfrac{1}{2}\pi r^2 \rho_a v^2 \sec\chi\, dz, \quad \ldots \quad \ldots \quad (19)$$

where r—radius of the meteor,

ρ_m—density of the meteor material,

l—latent heat of evaporation of the meteor material,

v—velocity of the meteor,

ρ_a—density of the air,

z—height of the atmosphere reckoned from some fixed level,

χ—angle between the meteor track and the vertical.

Since the momentum is conserved, we write

$$\tfrac{4}{3}\pi r^3 \rho_m dv = \pi r^2 \rho_a v \sec\chi\, dz. \quad \ldots \quad \ldots \quad \ldots \quad (20)$$

For the short path of the orbit of the meteor under consideration, the zenith angle χ may be assumed to be constant.

Assuming that the atmosphere is in hydrostatic equilibrium, Eqs. (19) and (20) yield

$$r/r_\infty = \exp\left[-(v_\infty^2 - v^2)/(12l)\right] \quad \ldots \quad \ldots \quad \ldots \quad (21)$$

$$\frac{p \sec\chi}{r_\infty} = \frac{2g\rho_m}{3} \exp\left(-\frac{v_\infty^2}{12l}\right)\left[Ei\left(\frac{v_\infty^2}{12l}\right) - Ei\left(\frac{v^2}{12l}\right)\right] \quad \ldots \quad (22)$$

where r_∞ and v_∞—radius and velocity of the meteor outside the atmosphere,

$Ei(x)$—integral logarithm,

p—atmospheric pressure.

According to Eq. (22), evaporation begins as the meteor enters into the atmosphere.

If n be the number of atoms evaporated in unit time, then

$$n = -4\pi r^2 \rho_m (dr/dt)(N/A) \quad \ldots \quad \ldots \quad \ldots \quad (23)$$

where A—atomic weight,

N—Avogadro's number.

From Eqs. (19) and (23) we obtain

$$\frac{n \sec \chi}{r_\infty^3} = \frac{\pi N}{2gHAl}\left(\frac{p \sec \chi}{r_\infty}\right)\left(\frac{r}{r_\infty}\right)^2 v^3 \quad \ldots \quad \ldots \quad (24)$$

where H is the scale height.

From the above equations we can determine how the velocity v, the radius r and the number of atoms evaporated per sec. n, vary with the pressure p of the atmosphere for a meteor with a given value of the radius at entry (r_∞), a given inclination of the trajectory (χ) and a given initial

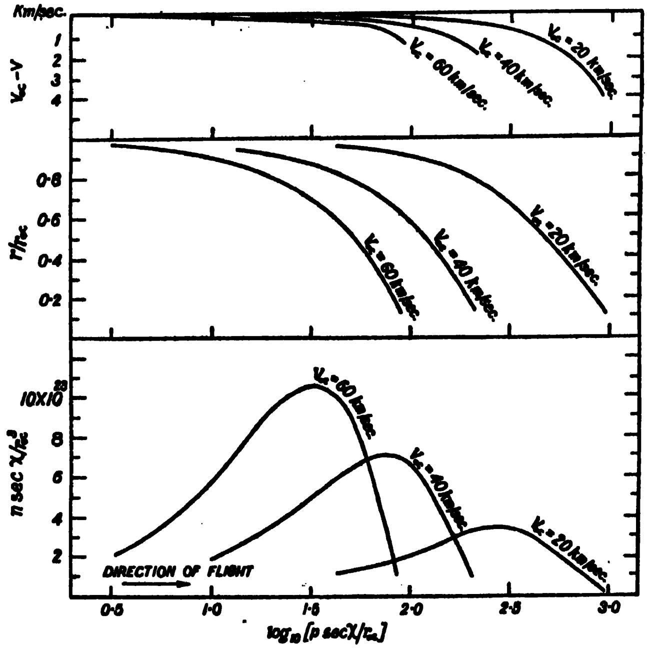


Fig. 9. Curves showing how the velocity, the radius and the number of atoms evaporated per second vary along the trajectory of a meteor. Three cases, with initial velocities 20 km./sec., 40 km./sec., and 60 km./sec. are illustrated. (After Herlofson.)

velocity (v_∞). In Fig. 9 is plotted $(v_\infty - v)$, r/r_∞, $n \sec \chi / r_\infty$ against $\log_{10}[p(\sec \chi / r_\infty)]$ for three initial velocities $v_\infty = 20$ km./sec., 40 km./sec. and 60 km./sec. The velocity 40 km./sec. is nearly equal to the critical velocity of the meteoric orbit round the sun.

We note the following interesting points from the three sets of curves. Firstly, the velocity is very little affected within the length of the track

under consideration. Secondly, meteors with higher velocity, irrespective of their size (as long as the radius is much smaller than the mean free path), disappear at greater heights. Thirdly, the last set of curves shows not only the variation of the rate of evaporation with height but also variation of ionization density with height, because, the rate of evaporation n is proportional to the production of heat, radiation and ionization along the entire length of the track (taking the velocity within the length of the track as nearly constant).

The shapes of the theoretical curves as shown in Fig. 9 are in general agreement with the observed distribution of luminosity of a meteor track [20, 21]. Further, the values of the scale height (H) as deduced from the lengths of meteoric tracks agree with the values deduced from other observations. One is thus justified in concluding that the theory of the rate of evaporation as sketched above yields results of right order.

From the equations deduced above we can obtain with good accuracy the maximum rate of evaporation as also the values of the pressure $p_{\max}$, the radius $r_{\max}$ and the velocity $v_{\max}$ at the point where the rate of evaporation is maximum. Thus, if we replace $E_i(x)$ in Eq. (22) by e^x/x and take account of the fact that the velocity variation is small, then from Eq. (22) we have

$$n_{\max} \doteqdot \tfrac{4}{9}\,(N/HA)(\tfrac{4}{3}\,\pi r_{\infty}^{3}\,\rho_{\max}\,v_{\infty}\cos\chi). \qquad \ldots \quad (25)$$

$$p_{\max} \doteqdot \tfrac{8}{3}\,lg\,\rho_{m}\left(r_{\infty}/v_{\infty}^{2}\right)\cos\chi \qquad \ldots \quad \ldots \quad \ldots \quad (26)$$

$$r_{\max} \doteqdot \tfrac{2}{3}\,r_{\infty} \qquad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (27)$$

$$v_{\infty}-v_{\max} \doteqdot \frac{6l\log na \pm 1{\cdot}5}{v_{\infty}}. \qquad \ldots \quad \ldots \quad \ldots \quad (28)$$

According to Eq. (25), the maximum rate of evaporation is proportional to the vertical component of the initial momentum of the meteor.

We further note that under the above condition the rate of evaporation at any point of the track is given by

$$n \doteqdot \tfrac{9}{4}\,n_{\max}\,(p/p_{\max})\,\{1-\tfrac{1}{3}\,(p/p_{\max})\}^{2}. \qquad \ldots \quad \ldots \quad (29)$$

This last equation shows the interesting fact that the shape of the evaporation curve does not depend on the initial size or the initial velocity of the meteor.

Let us now consider how the evaporated atoms will excite luminous radiation and also produce ionization.

As already mentioned the evaporated atoms will have, with respect to the atmospheric molecules, practically the same velocity as that of the meteor from which they originate. If the meteor atoms be assumed to be those of iron, then, for typical meteor velocities, this energy will vary from 100 eV for 20 km./sec. to 1,000 eV for 60 km./sec. Collision of iron atoms of such energy values may be expected to convert their kinetic energy into heat, light and ionization. Unfortunately, little data are available for the transition probabilities for collision phenomena with heavy particles, particularly in the above energy range. It is however known that the

probabilities of transition for such collisions are extremely small compared to probabilities for collisions with electrons. We recall in this connection that for collision with electron the probability of transition is maximum when the energy of the electron is a little above the energy of excitation or ionization. Recent experiments with heavy particles show that probabilities comparable to above are reached only when the kinetic energy exceeds 10,000 eV. It may be mentioned that this difference in behaviour between collision with electron and collision with heavy particle is due to the difference in mass. This necessitates, for the latter process, transfer of a much larger amount of momentum for transforming a given amount of kinetic energy into energy of excitation. It may be noted that Öpik [22] from his theory of atomic collision has deduced the probabilities of such excitation and has obtained the order of magnitude as indicated above. According to Herlofson, however, the probability of ionization as deduced above is rather high. Considering everything, Herlofson estimates that the kinetic energy of a typical meteor will be divided in the ratio $10^4 : 10^2 : 1$ for production of heat, light and ionization respectively.

Starting with the above assumption we can make an estimate of the number of electrons produced per cm. path for a meteor of given characteristics. Thus, from Eq. (24) or Fig. 9 we find that for a meteor of velocity 40 km./sec.,

$$n_{max} = 7\times 10^{23} r_{\infty}^3. \qquad \text{(30)}$$

Hence, the maximum number of electrons produced per cm. path of the track is

$$\alpha_{max} = 10^{-2}\, n_{max}/v = 2\times 10^{15} r_{\infty}^3. \qquad \text{(31)}$$

The maximum energy emitted as light per unit time,

$$I = \tfrac{1}{2} n_{max}(A/N)v^2\times 10^{-2}$$

$$= 5\times 10^{12} r_{\infty}^3 \text{ erg/sec.} \qquad \text{(32)}$$

According to Öpik [23] an object at a distance of 100 km. emitting light of intensity I will have the stellar magnitude,

$$m = 24{\cdot}6 - 2{\cdot}5 \log I. \qquad \text{(33)}$$

Hence, for the intensity I as given in Eq. (32)

$$m = -7{\cdot}2 - 7{\cdot}5 \log r_{\infty}. \qquad \text{(34)}$$

We can now construct the following table for meteors of two typical magnitudes +1 and +6, moving with velocity 40 km./sec., with the help of Eqs. (26), (31) and (34):

Stellar magnitude	+1	+6
Radius	0·08 cm.	0·02 cm.
Number of electrons per cm. path	10^{12}	10^{10}
Height corresponding to maximum of light and ionization	85 km.	95 km.

In Fig. 10 light curves—absolute magnitude plotted against time—of three meteors, as photographed by the Harvard group (Luigi G. Jacchia).

are shown [23a]. The smooth curves are the theoretically computed curves. It will be noticed that though the general trend of the actual light curves is the same as the theoretical curves of evaporation in Fig. 9, there are irregular fluctuations of intensity (meteor Nos. 808 and 1544). In the lowermost curve (meteor No. 1587) there is a 'burst' or 'flare'. This has cut short the life of the meteor. The origin of the irregularities in the light curves and of the flares is still unexplained.

The light curves of the meteors in Fig. 11 show 'humps' when they pass through the E-layer of the ionosphere. The phenomenon is extremely interesting and demands further study and clarification.

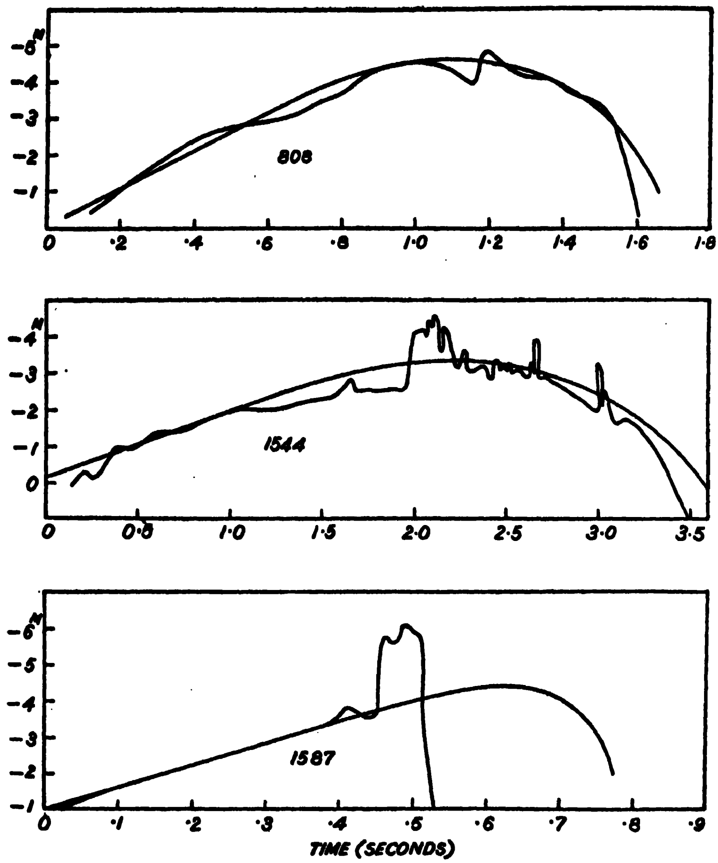


Fig. 10. Light curves of three meteors. Absolute magnitude of the meteor is plotted against time. (The numbers against the curves indicate the serial number of the observed meteor.) The smooth curves are the theoretically computed curves. The general trend of the light curves is the same as that of the evaporation curves shown in Fig. 9. There are, however, irregular fluctuations of intensity. In the lowermost curve there is a burst or flare. This has cut short the life of the meteor. [After Jacchia; Technical Report No. 3, Harvard College Observatory and Centre of Analysis of M.I.T. (1949).]

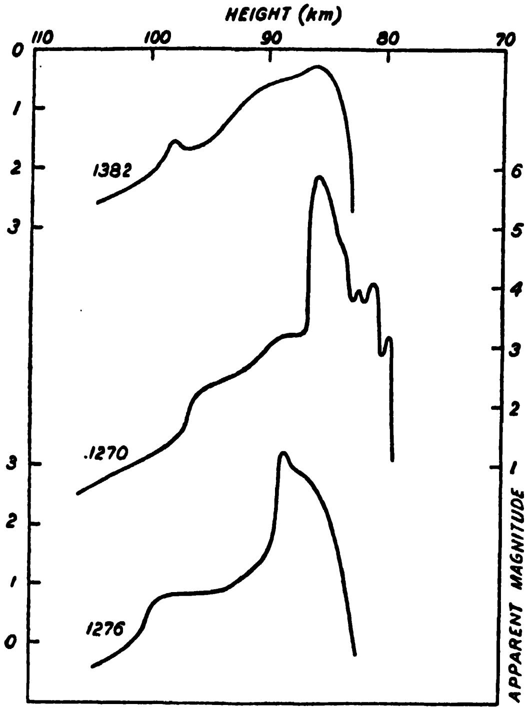


FIG. 11. Light curves of these meteors showed increase in brilliancy when crossing the *E*-layer. [After Jacchia; Technical Report No. 3, Harvard College Observatory and Centre of Analysis of M.I.T. (1949).]

(*d*) F. L. Whipple's computation of density and temperature

F. L. Whipple has utilized the photographic meteor data as recorded during 1936-37 by the Harvard group (*vide supra*) for computation of density and temperature distribution in the middle atmosphere [9]. The theory adopted by him is based on contributions by Öpik [22] and by Hoppe [24]. Formulae have been developed for calculation of density at four significant points along a meteor trail, namely, near the beginning, at the maximum light, at the end, and at any point where the deceleration is measurable. Results as obtained for these four points are generally consistent amongst themselves, the systematic deviation being 0·1 in log (density). Depending upon the weighting ascribed to the different methods of calculating the density, a considerable range of solutions occur. The best solution is found to be one in which the height-log(density) curve corresponds to the following temperature distribution: A flat temperature maximum of about 375°K at the 60 km. level, a rapid drop to 250°K near 80 km., and a constant or slowly rising temperature at greater heights to about 110 km. A seasonal effect is indicated. The upper atmosphere is raised 5·3±1·0 km. under average midsummer temperature as compared to its height under average midwinter temperature.

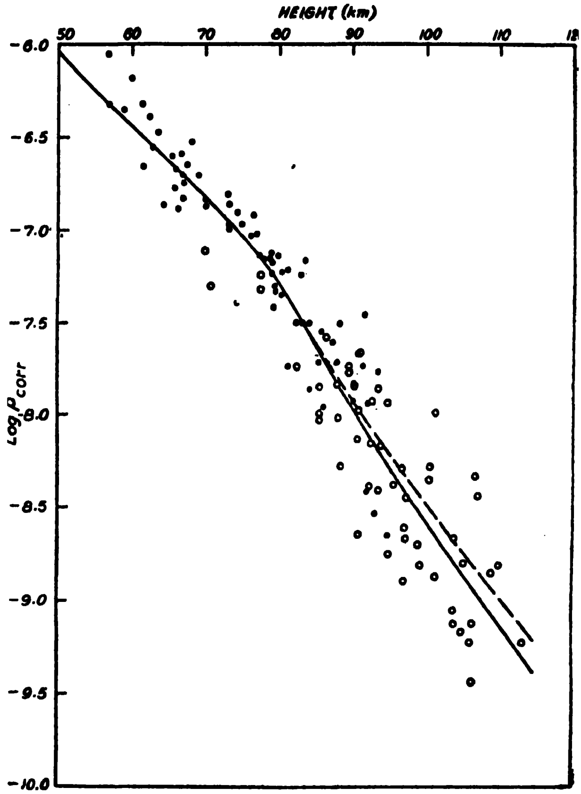


FIG. 12. Atmospheric density profile from meteor trails. Open circles represent density computed from magnitudes, velocities and masses of the meteors near the beginning of the photographic trails (corrected for velocity and seasonal effect). Dots represent density from meteor deceleration equally corrected. The upper curve shows the mean density variation with height. The lower curve follows the standard density variation as adopted by National Advisory Committee in Aeronautics of the U.S.A. [After Jacchia; Technical Report No. 4, Harvard College Observatory and Centre of Analysis of M.I.T. (1949).]

In Fig. 12 results of computation of atmospheric densities are plotted after Jacchia of the Harvard group [24a]. The open circles have been obtained from the magnitudes, velocities and masses of meteors near the beginnings of their photographic trails; the dots are from decelerations. These are all corrected for velocities and seasonal effects. The mean density curve (the upper curve) is drawn through these plots. It is to be noted that these determinations agree well with those of Whipple given above. There is, however, a systematic difference with V-2 rocket measurement. (See Chapter XII.) Above 65 km. height the meteor results give progressively greater densities than V-2 measurement results. It appears that meteor and rocket measures of density can only be reconciled by making fundamental changes in the form of the meteor equations [10].

5. METEOR OBSERVATIONS—RADAR STUDY

Note—Some of the terms used and experimental equipment referred to in this section will be better understood if Sec. 4 of Chapter VI on the Ionosphere is read first.

(*a*) Introduction

In Sec. 3 we have described the visual and photographic methods of meteor observation. Such observations are necessarily confined to dark clear nights only. In the present section we will describe the new and powerful radio-echo (or the radar) technique of meteor observation which does not suffer from these limitations. Meteor study by this method is possible in full daylight and through overcast sky.

This new method has been made possible by the intense ionization produced along meteor trails, to which we have already referred in the last section. It might also be mentioned in this connection that the possibility of meteor trails producing ionization had been suggested by many of the early ionospheric investigators [25, 26, 27, 28, 29]. Some workers had also obtained distinct correlation between sudden and transient radio echoes with passages of single meteors overhead [30]. It is, however, only in comparatively recent years that one has been able to form a clearer picture of meteor ionization and has also been able to make quantitative estimates, both theoretically and experimentally, of the ionization density in the trail.

Various considerations show that the ionization in meteor trails is initially in the form of long, thin columns some 10 km. in length, with electron density about 10^{10} to 10^{12} per centimetre length. The radius of the column is initially comparable to the mean free path in the region of the trail. This is of the order 10 cm. in the region of 100 km. The initial volume density of electrons in a meteor trail in this region is therefore 10^8 to 10^{10} per cm.^{3} With the passage of time the column spreads by diffusion and forms large patches or clouds of dense ionization. Evidences have now accumulated showing that the aggregate effect of these ion clouds is confined within a narrow limit of heights (altitude 80 km. to 120 km.) just below and partly overlapping the E-layer of the ionosphere. The meteor ionization has thus been found to reveal itself broadly by two types of radio 'echoes'. Individual meteor trails, immediately on their formation, produce radio echoes of a very transient character. These are best distinguished on wavelengths in the metre range. Also, the layer of patchy ionization—the aggregate of ion clouds—produces in the ordinary ionosphere equipment radio echoes which are more persistent than the former, on frequencies much higher than the E-layer penetration frequency. This second type of radio echoes will be discussed in Chapter VI, Sec. 13(*b*). In the present section we will discuss the first type of echoes. The study of this type has gained great impetus in recent years because of the developments during World War II in the design and operation of 'pulsed' transmitters and

corresponding receivers on frequencies much higher than the E- or F-layer critical frequencies. Active centres of systematic meteor study by the radar technique have grown up at a number of places, e.g. Jodrell Bank, University of Manchester, England [31], Stanford University, U.S.A. [32], and the Dominion Observatory and the National Research Council, Ottawa, Canada [33]. Mention should also be made here of the studies of meteor trails by the radar technique made in Japan [33a]. (Fig. 13, Plate III, is an aerial view of the Jodrell Bank Station.)

In what follows we will first discuss the intensity of the echo as expected from reflection of radio wave by meteor trail. The results of radar study of meteor ionization, e.g. heights of the meteor trails, positions of the radiants, velocities, occurrences of daylight meteor showers, will then be described.

The pulse technique, it may be recalled, was originally developed for exploring the ionized regions of the upper atmosphere—the ionosphere (see Chapter VI, Sec. 4). It employs a radio transmitter designed to send out groups (or 'pulses' as they are more usually called) of radio waves of very short duration—a few micro-seconds—in quick and regular succession. These are reflected when they meet ionized regions of appropriate electron density. The reflected wave or 'echo' is picked up by suitably designed receivers and its study yields valuable information regarding the ionosphere, e.g. heights of the ionized regions, their electron densities, and other characteristics. The frequency used in ionosphere 'pulse' equipment seldom exceeds 20 Mc./sec.

The pulse technique was later adapted for obtaining radio echoes from distant objects like aircraft. Thus has developed the elaborate system of equipment for radio-location—the Radar—by which the range, height and azimuth of a distant aircraft (or for the matter of that, any distant radio wave reflecting object) may be determined, in the dark and/or when the sky is overcast. Radar work is generally carried out with waves in the centimetre or metre range, with frequencies much higher than those employed for ionospheric work. Detailed description of radar technique is not given here as this is readily available in the many standard works on radar engineering. A 'pulse' equipment for ionospheric study will be described in Chapter VI, Sec. 4.

(*b*) Scattering of radio waves from meteor trails

To estimate the intensity of the radio echo from a meteor trail we can idealize a trail as having the shape of a long cylinder of intense ionization. The scattered intensity will then depend (besides upon the density of ionization) on the diameter of the column—whether it is large or small compared with the wavelength of the incident wave.

If the diameter is large compared to the wavelength (as assumed by Pierce [34]), then there will be total reflection from the body of the meteor at normal incidence if the electron density in the trail is greater than $\pi m f^2/e^2$ (f is the frequency of the incident wave and e and m have their usual significance. See Chapter VI, Sec. 2). But, as shown by Herlofson [16], this leads to impossibly high values of electron density.

If the diameter is small compared to the wavelength (as assumed by Blackett and Lovell [35]) there is scattering. The radius being small, the electrons within the same cross-section will scatter with the same phase and the thin column will not appreciably modify the incident wave.

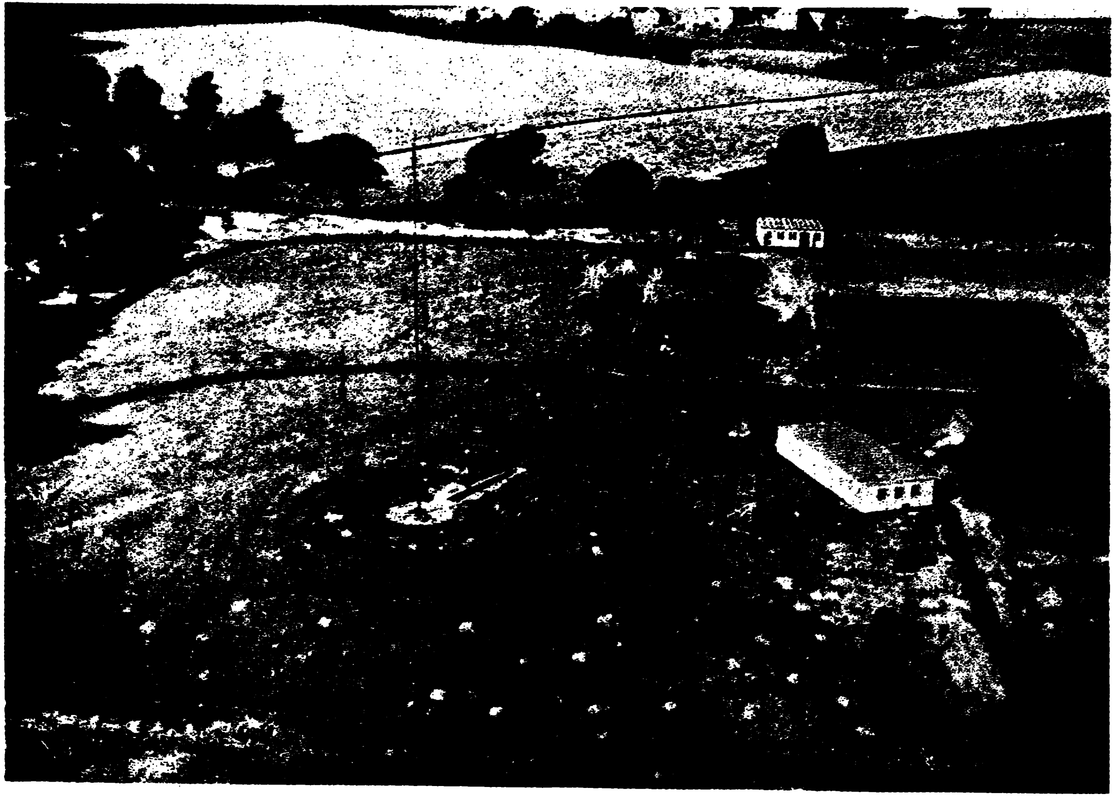


FIG. 13. Aerial view of the Jodrell Bank Station, University of Manchester, England, for meteor study by the radar technique.

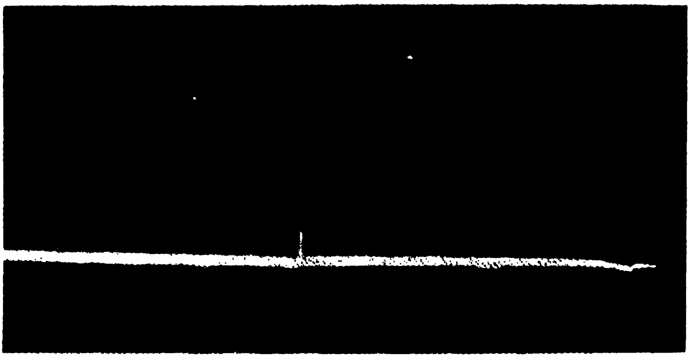


FIG. 15. Range-amplitude display of radio echo (broadside) from the trail of a meteor. The range (distance) of the 'reflecting' point of the trail and the strength of the echo are measured from the record. (After Lovell.)

FIG. 16. Intensity-modulation display of the radio echo (broadside) from a meteor trail. The duration of the echo and the range (distance) of the 'reflecting' point are measured from the record. (After Lovell.)

Electron density as calculated, on this assumption, from the observed echo intensity agrees with theoretical estimates.

To determine the intensity of the scattered echo on this assumption consider Fig. 14 in which XY is the trail of a meteor. Radio waves are

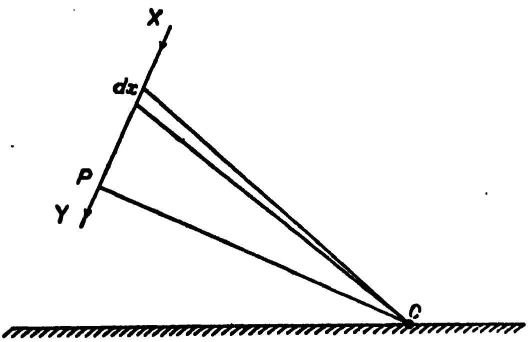


FIG. 14. XY is a meteor trail. Radio waves in the metre range are sent out from O and the echo scattered back from the trail is received also at O. The intensity of the received echo at O is appreciable only when there is a portion of the trail at right angles to the direction of observation.

sent out from O and the observer, situated also at O, receives the reflected echo. The intensity of the echo is determined by the effect of superposition of scattered wavelets sent out from elements like dx and received at O. We can, as in the treatment of diffraction problems in optics, divide up the whole length XY, starting from P, the foot of the perpendicular from O, into Fresnel 'zones', so that the mean distances from O of the successive zones differ by half-wavelength. It can then be shown that the resultant amplitude at O is determined by the effect of the first zone at P—the effects of the outer zones cancelling each other by mutual interference. The point which is nearest to O can thus be taken as the point of 'reflection' of the wave from the trail. It thus follows that radio echo will be obtained from a trail only when a portion of the trail is at right angles to the line of sight. This type of reflection may therefore be called 'broadside' reflection. (If the transmitter and the receiver are not at the same place as in Fig. 14 then the 'reflection' point is so situated that the incident and reflected rays make supplementary angles with length of the trail.)

We can thus assume after Lovell and Clegg [36] that, in accordance with the diffraction effect formula, only the electrons contained within column of length equal to $\sqrt{\lambda R/2}$ are effective in building up the reflected echo strength (λ—wavelength, R—range of the meteor). If α be the electron density of the trail, i.e., the number of electrons per unit length, then the appropriate number N of electrons is given by, $N = \alpha\sqrt{\lambda R/2}$. (*Note:* This formula differs from the more familiar formula $N = \alpha\sqrt{\lambda R}$, because the phase correction is introduced twice, on account of the curvature of the incident and reflected wave surfaces.)

Expressions convenient for numerical calculations from observed radio data may now be written after Lovell and Clegg as follows:—

$$\epsilon = \alpha^2 \frac{PG^2\lambda^3}{24\pi^2 R^3}\left(\frac{e^2}{mc^2}\right)^2 \text{ watts.}$$

Or, alternatively, $$\alpha = \sqrt{24}\,\pi\frac{mc^2}{e^2}\frac{1}{G}\sqrt{\left(\frac{\epsilon R^3}{P\lambda^3}\right)} \quad \ldots \quad \ldots \quad \ldots \quad (35)$$

where, ϵ—energy at the receiver input due to scattering from a meteor trail with α electrons per cm. length,

P—peak power in the pulse transmitter in watts,

G—power gain of the receiving and the transmitting aerials over a half-wave dipole, and e, m and c have their usual meanings.

The above expression for scattered intensity is applicable only to cases where the electron density is less than the critical density for total reflection —a condition satisfied by the majority of meteor trails even up to wavelength of 4 m. But, there may be exceptionally brilliant meteors for which the electron density in the core is higher than the critical density even for such high frequencies. For such cases, the intensity of scattering depends upon the angle between the length of the trail and the direction of the electric vector in the wave. Detailed theory shows that in such case the intensity will be greater if the electric vector is normal to the electron trail [38, 41].

Besides the broadside echoes discussed above, there is another type of echo due to, what may be called, end-on reflection. The electron density in the head of a trail, as it grows, is sometimes so high as to exceed the critical density for reflection of 70 Mc./s. waves. As such, waves of this frequency occasionally (from exceptionally bright meteors) and of lower frequency more frequently are reflected from the head. Use has been made of echoes by this type of reflection for measuring meteor velocities and will be referred to again.

Figs. 15, 16, Plate IV, are photographs of broadside radio echoes from trails as seen on the cathode ray oscillograph screen by conventional methods of display. Fig. 15 is range-amplitude display, in which the range (distance) of the 'reflecting' point of the trail and the strength of the echo can be measured. Fig. 16 is intensity-modulation display in which the duration of the echo and the range of reflecting point of the trail can be measured.

(c) Electron density in the trail

Eq. (35) can be readily used for estimating α, the number of electrons per unit length of the column, since, for a given meteor trail, ϵ and R can be measured, and G, P and λ are known constants of the radar apparatus. Such measurements have been carried out by Lovell and Clegg [36] and it has been found that the density in the trail of a 5th magnitude meteor (which is on the limit of visibility) is approximately 2×10^{10} electrons per cm. length. For meteors of magnitude about $+1$, the density is of the order 10^{12} electrons per cm. length. These results are in very good

agreement with the theoretical estimates of Herlofson [16] made on the basis of Öpik's theory of meteors. (See Sec. 4(*c*).)

(*d*) Life of a meteor trail

As already mentioned, the meteor trails are initially in the form of long thin column of ionization, some 10 km. in length and diameter of the order 20 cm. (at 100 km. height). Though the electrostatic forces between the electrons and the positive ions tend to retain the structure, there are forces—turbulence, diffusion and recombination which ultimately destroy it. If the effect of the terrestrial magnetic field on the motions of the electrons and ions is neglected for the moment, then the rate of decrease in ionization density with time may be estimated without difficulty, by making reasonable assumptions regarding recombination and diffusion rates. According to Lovell [39] if these two simple processes are only operative, the duration of a radio echo from a meteor trail should be from a fraction of a second to a second or so. Further, for a given trail, the duration of the echo, for the same trail, will be longer with longer wavelength. Also, the duration will decrease rapidly with increasing height of the trail, due to increased diffusion rate.

Of these deductions, the second and third accord well with observations. Simultaneous observations of the same trail, on two wavelengths, with apparatus of equivalent sensitivities show the expected increase in duration with increase in wavelength. It has also been found that the duration decreases with increase in velocities. Since meteors of high velocity appear and disappear at higher altitude, this, in other words, means that the duration of the echo decreases with altitude.

In regard to the average duration of the echo the observational results are confusing. For example, in the case of a given meteor shower of homogeneous velocity, the duration has been found to vary from less than 0·1 sec. to greater than 60 sec. Also, in the set of observations on relation between altitude and duration referred to above, it was found that in some 10% of cases the echo could be observed for many seconds, that is, some hundreds of times greater than the expected value. These abnormally long duration echoes also show complex and violent fluctuations in intensity. It has been suggested that these fluctuations may be due to turbulence [40]. Many workers have also observed that the range of these long duration echoes drift with time. These anomalies, as also the main problem, namely, the occurrence of echoes of unusually long duration—which means that a sufficient density of electrons in the trail is maintained for periods of a minute against the normal forces of diffusion—are still unexplained. It should be mentioned here that besides ionization, as discussed above (which may be called the *basic* ionization), there is evidence that additional *secondary* ionization is produced by strong ultraviolet radiation around the meteor itself [39*a*]. This ionization is much more attenuated than the basic ionization and disappears rapidly within the time expected.

Radar echoes from the secondary ionization were first observed by Hey, Parsons and Stewart [39*b*].

In considering diffusion, the effect of the earth's magnetic field has been neglected. If this is taken into account then the combined diffusion of the ions and electrons will be slower in the direction normal to the magnetic field than along the same [38]. The ionized matter will thus spread into an elliptic distribution round the track of the meteor with the minor axis normal to the direction of the earth's magnetic field and the track of the meteor. There will, of course, be no distortion from axial symmetry for tracks lying along a magnetic line of force. There is experimental evidence to show that the direction of formation of the trail relating to the direction of the terrestrial magnetic lines of force has considerable influence on the duration and intensity of an echo.

(*e*) Variation of intensity and of frequency of the echoes with change of wavelength

Since the voltage amplitude (V) at the receiver input is given by $\sqrt{r\epsilon}$, r being the input resistance of the receiver, it follows from Eq. (35) that V varies as $\lambda^{3/2}$. For the same meteor, therefore, the variation of echo strength with change of wavelength will follow this law. This check on the formula has been carried out by Lovell and Clegg [36]. The agreement in the range of wavelengths 1·4 m. to 8·3 m. is quite good.

Apart from the variation of intensity of the echoes with wavelength, it has been found that the number of echoes depends very critically on the wavelength. During periods of no-shower, the number of echoes obtained on wavelengths below 6 m. is very few, and those that are observed have a high correlation with meteors that are observed visually. On wavelengths above 8 m., however, during these same periods and with apparatus of equivalent sensitivity, the number of echoes is enormously increased. They produce an extremely high background in which the visual type of echoes are submerged.

The increase in the echo-rate at 6 m. to 8 m. has been sought to be explained by the plasma resonance theory of ionized columns developed by Herlofson [41]. According to this theory the meteor trail behaves as a resonator for the longer wavelengths (when the value of the dielectric constant is near -1) and the echo strength becomes dependent on the angle between the electric vector of the incident wave and the axis of the trail. If the two are at right angles strong echoes are produced. The effect is ascribed to space charges induced in the boundaries of the trail as in plasma resonance studied by Tonks and Langmuir [42]. As the effect is more pronounced on longer wavelengths, the increase in echo-rate at 6 m. to 8 m. is explained. It may be mentioned that the dependence of the scattered intensity on the direction of the electric vector relative to the length of the column has been verified by laboratory experiments of Romell [42*a*].

It has also been found by Millman and McKinley [33] that on the lower frequencies, reflections may be obtained from a trail even when it is

not at right angle to the line of 'sight'. Further, according to these authors, the initial column of uniform ionization changes rapidly to one of localized concentrations of electrons. Hence, while on higher frequency only the 'right angle' reflection is observed, on lower frequencies a multiplication of echoes may be obtained. These phenomena also help to explain the increase in echo rate when the exploring wave frequency is lowered from 72 Mc./s. to 36 Mc./s.

(*f*) Measurements of height and range

The outline of a method for measurement of the height and range of meteor trails from observations made at a single observing station as developed by Clegg and Davidson [43] is described below. The method is based on the spaced aerial method for height determination as used in radio location work [44].

The region of the sky to be explored for meteors is flooded with 'pulsed' radiation from a suitably directed transmitting antenna. Two receiving aerials, R_1 and R_2 mounted at different heights above a reflecting surface pick up reflected waves from meteor trails (see Fig. 17). Each of the two

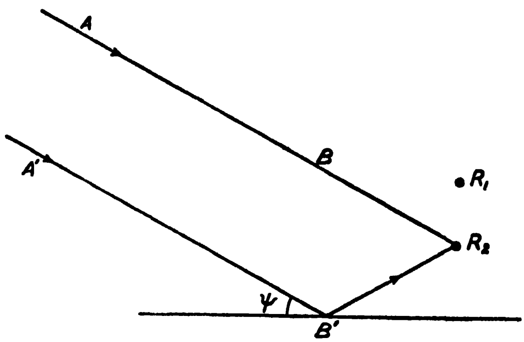


FIG. 17. Illustrating the method of estimating the angle of elevation of a meteor trail from the scattered signal received simultaneously by two aerials, one situated above the other over a reflecting surface.

aerials receives the waves sent out from the meteor trail (at an angle of elevation ψ above the ground) by two paths—one direct and another indirect *via* the reflecting surface. (In the figure the two paths ABR_2 and $A'B'R_2$ are shown for the aerial R_2.) The signal strength in each of the aerials is thus determined by the superposition effect of these two sets of waves. But, as the phase differences between the direct and the indirect waves are different for the two aerials, the signal strengths are also different. And, since the phase difference depends on the angle of elevation, the ratio of the signal strengths from the two aerials is a function of ψ. This function is easily calculable and hence the elevation of the reflecting point of the meteor trail can be determined from a measurement of the signal ratio. In practice, on account of different directive properties, the gains of the two aerials vary with the direction

ψ in different manners. But as these are calculable from known parameters of the system, the relationship between (r) the ratio of the effective gains $G_1(\psi)$ and $G_2(\psi)$ of the two aerials and the elevation angle ψ is easily established.

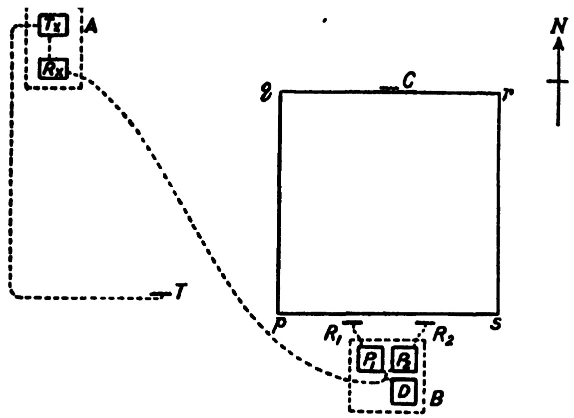


FIG. 18. Disposition of transmitter, receivers and transmitting and receiving aerials for measuring the angle of elevation of a meteor trail. *pqrs* is ground surface covered by wire mesh. (After Cleggs and Davidson.)

Fig. 18 shows the layout of the arrangement. The hut A houses the pulse transmitter T_X and the receiver R_X. These are of conventional design as used in radar work. The transmitter wave frequency is 60 Mc./s. ($\lambda = 5$ m.). The pulse duration is 8 microsec. and the peak power 100 kW. The receiver amplification is linear and its sensitivity is such that 2μ V signal input doubles the signal to noise ratio. The transmitting aerial T consists of a horizontal folded half-wave aerial at a height of $\frac{3}{8}\lambda$ above the ground, with a reflector to give minimum backward radiation. It floodlights the northern quadrisphere of the sky.

The two receiving aerials R_1 and R_2 are similar to the transmitting aerial. The distance between them and the hut A is about 200 ft. The two aerials are mounted on posts 20 ft. apart and at heights, normally $\frac{1}{2}\lambda$ and $\frac{3}{4}\lambda$. In front of the aerials is a plane macadamized square *pqrs* with sides 100 ft. in length. It is covered with wire mesh to form a reflecting surface for the received waves. (See Fig. 13; the black rectangular patch at top right.). The hut B contains two pre-amplifiers P_1 and P_2 to which the receiving aerials R_1 and R_2 are connected. The signals pass from here to the receiver R_X through a common co-axial cable. The pre-amplifiers are suppressed alternately at a frequency of 25 cyc./sec. by multivibrator pulses sent out from the hut A. The same multivibrator pulses are also applied to the X-plates of the cathode ray oscilloscope connected to the output of the receiver. The receiver being energized alternately by the two aerials R_1 and R_2 displays the echoes from the two aerials side by side and allows visual comparison of the relative amplitudes. (The fourth aerial

C is mounted for purposes of calibration to equalize the gains of the two pre-amplifiers P_1 and P_2 by observing the signals on the duplicate oscilloscope D connected to R.)

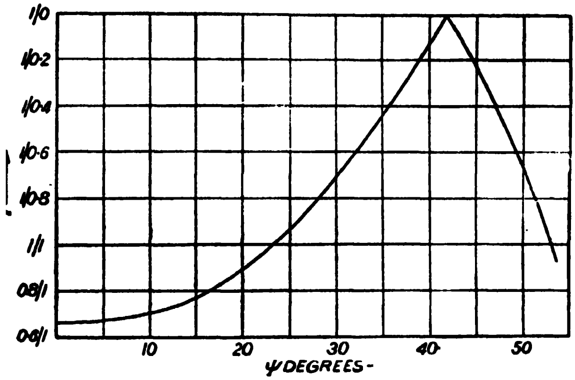


FIG. 19. Variation of the ratio r of the effective gains of the two aerials R_1 and R_2 of Fig. 18 with elevation. (After Clegg and Davidson.)

Fig. 19 shows the theoretical relation between the ratio of the effective gains of the two aerials R_1 and R_2 and the elevation angle ψ.

The aerials receive waves scattered back from all parts of the ionized meteor trail. But, as already explained, it is only the part of trail nearest to the observer—the 'reflecting' point, that is effective in building up the received signal strength. Hence, the amplitude of the received echo may be taken as proportional to the power gain in the direction ψ of the 'reflecting point' for either aerial and this direction is obtained by reading the graph.

Simultaneously with the elevation, the range is also measured from the time-delay of the received echo. From these two data the height of the 'reflecting' point of the trail above the surface of the earth is easily calculated.

The authors have discussed the limitations of and the errors in the method employed by them. The greatest uncertainty is in the estimation of the ratio r. The consequent error in estimation of ψ and hence of the height is ± 4 per cent. The error in range measurement is ± 1 km. Errors due to the imperfection of the reflecting surface and to the coupling of the two aerials are negligible.

The equipment has been used for measuring the heights of the 'reflecting point' of the meteors in the Quadrantids shower of 1949, January 3-4. The height distribution as obtained agreed well with the distributions of the heights of appearance and disappearance determined visually. The heights of the daytime showers of 1949 were also measured (see sub-sec. (h)). There is naturally no means of checking these latter results with visual observations. But the height distribution, as obtained, is the same as expected for meteors of medium velocity.

(g) Determination of meteor radiants

The property of a meteor trail giving broadside 'reflection', coupled with the fact that the meteor radiant moves across the sky with the rotation of the earth has been utilized for determining the position of the radiant on the celestial sphere. We describe here the main principles of two methods that have been developed for such determination. For fuller accounts the reader is referred to the original papers.

The first determination of meteor radiants was made by Hey and Stewart [37]. They employed three observing stations situated at three corners of a triangle—about 100 km. apart (South-East England). The direction of the aerial beam at each of the stations was so adjusted that the beams intersected at a height of about 100 km. above the ground. As the radiant moved across the sky, the meteor trails produced echoes at the three stations, not simultaneously, but in succession as the trails became oriented at right angles to the respective aerial beams. From the diurnal variations of the mean hourly rates of the echoes at the three stations, the position of the radiant could be approximately determined.

A more accurate method of determining the radiant is, however, due to Clegg [45]. The method is also more convenient as only a single observing station is required.

For determining the right ascension of a radiant the transmitted beam is pointed towards the east. Then, since the meteor trails must pass at right angle to the beam to produce the echoes, the echoes appear only when the radiant has moved to a position due south. Now, as already mentioned, the majority of the trails occurs within a limited range of heights round 100 km. (sub-sec. *a*). This limited zone intersects with the aerial beam. From a consideration of the geometry of this intersection it can be shown that the frequency of the echoes will be highest at long range and that as, with the rotation of the earth, the radiant moves towards the west the range of the echoes will steadily decrease. These variations in the echo rate and range are observed experimentally. Hence by noting the time at which the long range echoes first appear, the time of transit of the radiant is obtained and the R.A. calculated therefrom.

For determining the declination of the radiant, the aerial beam is now rotated through some angle θ towards the south. The beam is thus again in advance of the radiant. The echoes again appear at long range as the radiant moves into the position at right angles to the new direction of the beam. Then, as the radiant moves still further to the west, the ranges of the echoes decrease. One thus has the times of the first appearances of long range echoes at two positions separated by the angle θ. The declination is now calculated from this time interval and the angle θ through which the beam has been swung.

The accuracy of the method depends on the rate at which the echoes are received, on the width and the elevation of the aerial beam and on the declination of the radiant. It has been found that the changes in the range

and in the rate of occurrences of the echoes become very marked when the beam is narrow horizontally and is directed at sufficiently low elevation. With an aerial system whose half-amplitude beam width is $\pm 10°$, the probable errors in right ascension and declination are $\pm 2°$ and $\pm 3°$ respectively.

(*h*) Daylight meteor showers

Visual observations of meteoric phenomena are necessarily confined only to dark and clear nights. Observations by the radio method, however, do not suffer from these limitations. Meteor echoes may be obtained through overcast sky and in daylight hours.

Radio echoes on 4·2 m. observed during daylight by the Manchester (Jodrell Bank) group of workers from 1947 onward have been interpreted, after careful consideration, as due to a series of meteor streams [37, 45, 46, 47, 48]. The radiant points, the durations and the hourly rates of these streams have been determined. They occur between May and August. Their radiant points lie in a narrow strip of the celestial sphere extending from 330° to 120° right ascension, situated a little to the north of the ecliptic. The streamers are very active with hourly rates which may be compared to those of Perseids and Geminids showers. There were several occasions when two or three of these showers were active at the same time.

The remarkable features of these daylight streams are their duration—more than three months, and constancy in the distance from the sun of the main radiant point. No satisfactory explanation of these phenomena has yet been offered.

(*i*) Measurement of meteor velocities

Three techniques of measuring the velocities of meteors by radio echoes have been developed. We describe briefly the principles of these techniques. For the details the reader should consult the original papers.

The direct and straightforward method of measuring velocity is to record the time-range curve of the end-on echo from the advancing head of a meteor trail. [See sub-sec. (*b*)]. Hey [49] had been the first to measure meteor velocity by this method. The technique was later developed by Hey, Parsons and Stewart [50]. These authors used waves on 70 Mc./s. and, as such, the end-on echoes were rather rare. Millman and McKinley using wave on 30 Mc./s. made a large number of time-range records of such waves [33; 51, 52]. Fig. 20, Plate V, is a reproduction of one of the records obtained by Millman and McKinley [52]. It is a rare record in which both the approach and regression of the head of the meteor trail are observed. The velocity of the meteor can be determined since the record enables the range to be measured against time.

The second method, developed by Davies and Ellyett, is very interesting and may be called the Fresnel zone method [53]. Briefly, the method is as follows.

Consider Fig. 14 where XY is the trail of a meteor. Let us suppose that the meteor trail is growing near the foot of the perpendicular P. We have already explained that the echo amplitude at the receiving point O is determined by the superposition of waves sent out from the successive Fresnel zones in the trail and that it is only the zones near P that are effective in determining the echo intensity. It is thus evident that as more and more zones grow near P the amplitude at O will fluctuate, increasing or decreasing according as an odd or an even number of zones come into operation. The fluctuations are, in fact, similar to those observed near the line separating the geometrical region of shadow from that of light in the optical diffraction pattern by a straight edge.

For measuring the velocity, this intensity fluctuation is recorded and the rate at which the maxima succeed one another is measured. From this and knowing also the distance of the trail and the wavelength employed, the rate of growth of the trail (that is, the velocity of the meteor) is calculated.

For obtaining these data it is necessary to record automatically the range, amplitude, duration and variations in the strength of the echo within the first few milliseconds of the short span of the life of the echo. Signal strength variations are to be measured at milliseconds intervals, and, this can be done satisfactorily only by recording the amplitudes of successive received pulses with pulse recurrence frequencies of the order 500 to 1,000 per second. Such high recurrence of pulses, however, introduces difficulties in range measurements. The meteor echoes are commonly observed at ranges of 900 km., and the time interval between transmitting a pulse and receiving the echo becomes several times the interval between the pulses. But, notwithstanding these difficulties, apparatus has been designed by which it has been possible to record the fluctuations of the echo strength with the growth of the trail (see Fig. 21 Plate V).

The technique was first applied to the determination of meteor velocities during the Geminid shower, December, 1947. The velocity of the main group in the shower was found to be 34·5 km./sec., with perhaps a secondary group at 39·3 km./sec. Velocities of meteors during daylight showers of 1948 have also been measured. It is to be noted however that for various reasons the Fresnel zone effect can be distinguished only in a small percentage (6·3 per cent) of the recorded echoes.

The third method which may be called the radio Doppler method was first envisaged by Chamanlal and Venkataraman [54] and has been developed in recent years by Manning and his collaborators at Stanford, U.S.A. [55].

If continuous wave is used for end-on reflection from the head of a meteor trail, then the frequency of the reflected wave, since the reflecting surface is in motion, shows a Doppler shift. The reflected wave if heterodyned with the incident wave (intensity suitably suppressed) produces a beat frequency lying usually within the audible range. A little consideration shows that if the direction of the trail does not coincide with

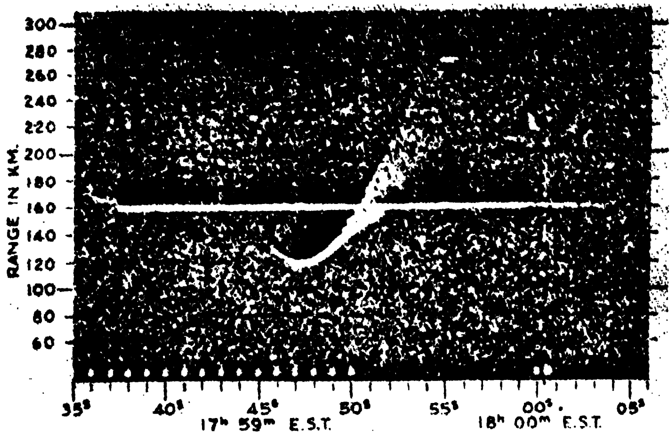


FIG. 20. End-on radio echo record from head of a meteor trail. Illustrates the approach and regression of the head. The velocity of the meteor can be determined from the record. (After Millman and McKinley.)

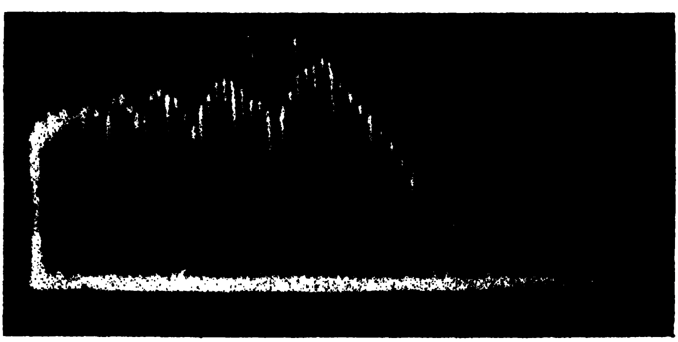


FIG. 21. Fluctuation of the echo strength with the growth of the meteor trail illustrating the 'Fresnel Zone' effect. (After Davies and Ellyett.)

the line of sight then, as the head of the trail advances, the component velocity of the head in the line of sight gradually decreases, becoming zero when the meteor trail is perpendicular to the line of sight [56]. The pitch of the beat note therefore decreases as the meteor advances, and produces a characteristic whistle which has been termed 'meteor whistle'.

These whistles were noticed even by early radio observers. In 1921 Pickard discovered that a sharp hiss could be heard in a highly sensitive receiver in exact coincidence with meteors [30]. Chamanlal and Venkataraman working at Delhi, India, reported in 1941 that weak heterodyne whistles were audible in a receiver tuned to the carrier wave on a nearby short wave transmitter radiating on any one of the frequencies 5, 7, 11 and 15 Mc./s. [54]. In most cases the whistle appeared as high pitched note which rapidly descended in pitch and faded away on reaching zero frequency or, sometimes before that. The duration of the whistles varied from approximately 1/5 second to several seconds. These whistles, the authors ascribed to beats between the direct wave and wave reflected from meteor with Doppler shift. They also attempted to make an estimate of the meteor velocity by assuming that the decrease in pitch was due to the velocity of the meteor being rapidly reduced. We know now, however, that this view is wrong. The meteor velocity remains sensibly constant and, as explained above, the descending pitch is due to the reduction of the component velocity in the line of sight [56].

As already mentioned Manning and his collaborators at Stanford, U.S.A., have made systematic study of the meteor whistles and have obtained meteor velocities therefrom. In one of their test experiments, the transmitter and the receiving system were 4 km. apart. Continuous waves of frequencies 23·1 Mc./s. and 30·66 Mc./s. were used. The radiated power was of the order of a kilowatt. The direct ground signal (with largely reduced intensity by suitable devices) was allowed to beat with the reflected meteor signals. Suitable recording devices were connected to the receiver output so that Doppler shift of frequency from zero to about 1,500 cycles/sec. could be recorded for study.

For determining the velocity it is necessary to know the distance of the meteor from the observing station. This was found by working simultaneously with a pulsed transmitter on 17 Mc./s. The pulses were of 150 micro-second duration with peak power of about a kilowatt.

If the rate at which the pitch is falling is known then a relation between this quantity and the velocity and range of the meteor is easily obtained. Thus,

$$v = (-f'_0 \lambda R/2)^{\frac{1}{2}},$$

where v is the velocity, f'_0 the slope with which the pitch approaches zero, λ the wavelength and R the range of the meteor. In the observations made very good plots of Doppler-pitch versus time have been obtained and the meteor velocities calculated therefrom.

The authors claim that the method is remarkably accurate and reliable, the deviations as observed for the set of measurements carried out during Perseid shower of August 11, 1948, being only a few per cent.

An interesting point regarding the times of detection of the 'broadside' and the 'end-on' reflection effects, first noted by Appleton and Naismith may be mentioned [56]. From Fig. 14 it is clear that the meteor whistle will commence even when the head of the trail is far away from P, the point of broadside reflection. The meteor whistle will therefore precede the broadside burst of echo. This is what has been observed by Manning *et al.* and by previous workers.

It should be mentioned, however, that there has been a different interpretation of the origin of the meteor whistle. According to McKinley [57], the majority of the 'Doppler whistles' are not due to Doppler shift at all, but are due to 'the intensity and frequency changes during the Fresnel zone formation and are, in fact, the audible continuous wave counterpart of the diffraction phenomena discussed above'. Real Doppler whistle is a rare phenomenon and is found only in cases of very intense ionization. It has also been found that there is little difference in the velocities (except near the point of broadside reflection) as derived from either assumption on the mode of production of the whistle.

CHAPTER IV

THE OZONOSPHERE

1. INTRODUCTION—HISTORICAL

Since the early days of spectroscopy, it has been known that the solar spectrum is cut off rather abruptly near 2900Å (Fig. 1). The abrupt ending is now known to be due to strong absorption by ozone in the atmosphere, the strongest part of which lies in the spectral region 2200Å to 2900Å. The ozone is spread in a diffuse layer between 10 and 50 km. above the surface of the earth with its centre of mass lying at about 25 km. The thickness of the layer reduced to S.T.P. is small, and variable—ranging from 1·5 mm. to 4·5 mm., averaging 2·5 mm. (The amount may be compared with that of such minor constituent of the atmosphere as carbon dioxide, the equivalent thickness of which at S.T.P. is 2·4 mm.) The ozone layer, or the *ozonosphere* as it is called, acts as a heat reservoir in the middle atmosphere and is responsible for the rise of temperature discussed in Chapters IIIA and IIIB.

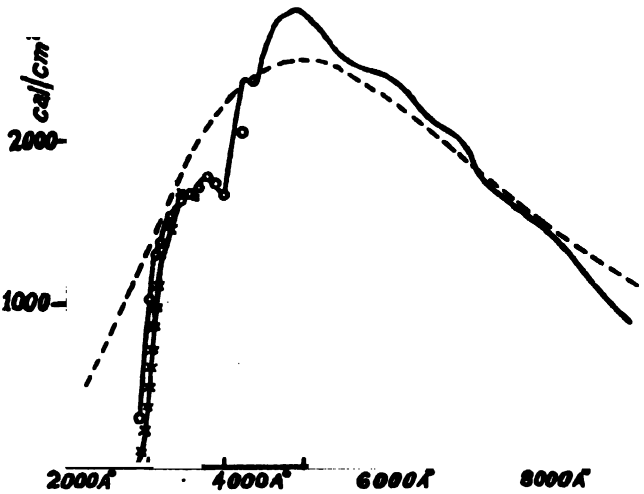


FIG. 1. Illustrating the abrupt ending of the solar spectrum near 2900Å (continuous line). The steep part of the continuous line curve (circles) is from measurements of Abbot [1] in the ultraviolet. The broken line curve shows the spectral distribution of energy for the sun at 6000°K. The crosses indicate the absorption by a layer of ozone 0·3 cm. thick. The close parallelism of the last named curve with the steep part of the solar curve shows that the sudden decrease of the energy in solar spectrum near 3000Å is due to ozone absorption. (After Abbot.)

Cornu in 1878 first suggested from his observations on the limitation of the solar spectrum towards the ultraviolet that this was due to absorption in the terrestrial atmosphere. The agency causing the absorption was suggested by Hartley [2] to be ozone when in 1880 he discovered the

absorption bands associated with his name, extending from 2100Å to 3200Å This view was strengthened in 1890 when the so-called Huggins bands were discovered in the spectrum of Sirius [3]. It was subsequently proved that these bands were due to absorption in the terrestrial atmosphere and that the particular gas responsible for their production was ozone. In 1917, experiments by Fowler and Strutt (Rayleigh) [4] gave the most conclusive evidence regarding the limitation of solar spectrum by ozone absorption. They compared the absorption spectrum of ozone with the spectrum of Sirius in the ultraviolet and proved definitely that atmospheric ozone is the effective agency in fixing the limit of solar spectrum towards shorter wavelength.

It was at first believed that the absorbing ozone was distributed in the atmosphere near the surface of the earth. Attempts were made, in fact, to extend the solar spectrum towards the ultraviolet by taking spectral photographs on mountain tops. It was soon discovered, however, that the solar spectrum could not be extended in this way. A natural conclusion of this was that the atmospheric ozone responsible for limiting the solar spectrum is situated in the high atmosphere. The first quantitative estimate of the height of the ozone layer was made by Strutt (Rayleigh) [5]. in 1917 from observations on the spectra of the rising and the setting sun. He concluded from his measurements that the ozone could not be distributed uniformly in the atmosphere and that the ozone layer lay between 40 and 60 km. above the sea-level. Later measurements by improved technique (*vide infra*) by Götz, Meetham and Dobson [6], showed that the atmospheric ozone is situated at a lower level and is distributed between 10 and 50 km. with a maximum at about 25 km. A brilliant confirmation of the existence of ozone in the upper atmosphere has been furnished by the balloon experiments of E. and V. H. Regener [7]. The sounding balloon carried a spectrograph with automatic recording arrangement. The highest ascent was up to 31 km. The registered spectra showed that with the rise of the balloon, as more and more of ozone was left below, the spectrum gradually extended towards the ultraviolet (Fig. 4).

2. ABSORPTION SPECTRUM OF OZONE

The strongest and the most important of the absorption bands of ozone is the Hartley band system in the ultraviolet responsible for the abrupt ending of the solar spectrum at 2900Å [2]. Underlying the Hartley bands there is an absorption continuum. Besides these bands in the ultraviolet, there are other band systems, much weaker, in the visible and in the infra-red regions. We give below short descriptions of the band systems in the different regions. For more detailed accounts the reader should consult references [8].

(a) Ultraviolet region

As mentioned above, the dominant band system in the ultraviolet is the Hartley band system extending from 2100Å to 3200Å. The bands

have a rather diffuse structure and appear against a background of continuous absorption. The latter has a high absorption coefficient ranging from 74 cm.$^{-1}$ at 2340Å to 15 cm.$^{-1}$ at 2130Å. The strong absorption by ozone was first measured by Fabry and Buisson [9 *a*] and later by Laüchli [9 *b*]. In Fig. 2 are depicted the absorption values obtained by Ny-Tsi-Zé and Choong-Shin-Piaw [10] and Chalonge and Lefebre [11]. It gives the variation of the absorption coefficient of the Hartley bands with wavelength and shows that the absorption is very strong.

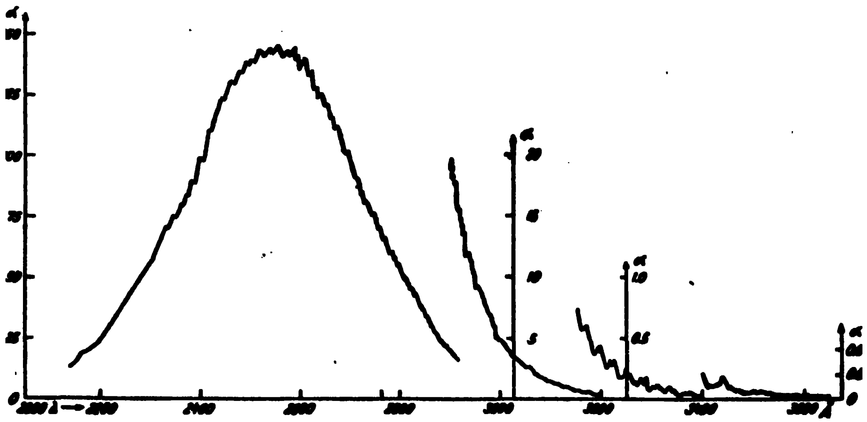


FIG. 2. Absorption by ozone in the ultraviolet. The strongest absorption is due to the Hartley bands. (Note that for the different regions, different ordinate scales have been used). The portion of the curve towards long wavelengths, 3400Å to 3600Å is drawn after Chalonge and Lefebre [11]. The rest of the curve is after Ny-Tsi-Zé and Choong-Shin-Piaw [10].

Overlapping the Hartley bands in the long wavelength side and extending beyond it are Shaver bands and Huggins bands [3]. Shaver bands overlap the Hartley bands on their long wavelength side. Huggins bands occupy the region 3200Å to 3600Å. Contrary to Hartley bands, the Huggins bands are sharp, the heads being towards the red. The absorption coefficient of these bands is very small compared to that of Hartley bands (Fig. 2).

(*b*) Visible region

The absorption of ozone in the visible region between 3600Å and 4500Å is extremely small. It appears, however, that this region has not been thoroughly investigated. Beyond this region comes the Chappuis diffuse band system extending from 4500Å to 6500Å. According to Colange [12], the absorption coefficient is very small, the maximum being only 0·05 cm.$^{-1}$ at 6100Å.

(*c*) Red and infra-red bands

The band systems in the red and infra-red have been studied by Maris [13] and also by Hettner, Pohlmann and Schumacher [14]. Three

band systems are known. The absorption maxima of these lie at $4{\cdot}7\mu$, $9{\cdot}6\mu$ and $14{\cdot}1\mu$ and the values of the absorption coefficients are 0·13, 0·22 and 0·70 cm.$^{-1}$ respectively.

(*d*) Effects of temperature and pressure on absorption bands

The effects of temperature and pressure change in the absorption band spectrum of ozone have been studied by various authors in the different spectral regions—infra-red, visible and ultraviolet. We give here briefly the results of study in the ultraviolet region as that is of the most interest to us. Wulf and Melvin [15] carried out observations in the range of −78°C. to 250°C. and found that when the temperature was lowered the Hartley bands between 3400Å and 2900Å exhibited strengthening in the maxima and weakening in the space between the bands. Vassy [16] has also investigated the temperature and pressure effects with great care between −80°C. and 20°C. and between 760 mm. and 19 mm. respectively. The pressure change had practically no effect. Contrary to the results of Wulf and Melvin, Vassy did not find any change in the maxima of absorption;

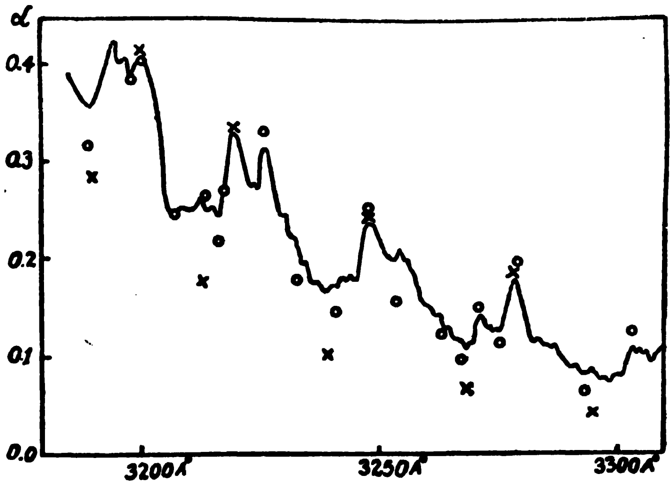


FIG. 3. Absorption by ozone at different temperatures. The continuous curve depicts the variation of ozone absorption at ordinary temperature (15°C.) within the wavelength range indicated (after Ny-Tsi-Zé [10]). The circles represent the absorption coefficients as obtained from observations on the blue sky. The crosses are for ozone at −80°C. (after Vassy [16]). It will be noticed that with the lowering of temperature the values of minima are lowered.

the minima, however, became more transparent particularly towards the long wavelength side (Fig. 3). The change was further found to be linear with the temperature. Inspection of Fig. 3 shows that the average temperature of the atmospheric ozone layer causing absorption should be much lower than the laboratory temperature (15°C.). This is as it should be, because, we shall presently see that the main concentration of ozone is in the low temperature region of the stratosphere.

3. VERTICAL DISTRIBUTION OF OZONE—EXPERIMENTAL METHODS

(a) Introduction

The distribution of ozone in the middle atmosphere can be obtained, (i) directly, with the help of sounding balloons carrying self-registering apparatus, and (ii) indirectly, from spectro-photometric study of the direct or scattered sunlight in the ultraviolet near the limit of the solar spectrum. The method of sounding balloons, though direct and reliable when the recording apparatus is properly designed, has been used only on rare occasions. This is because the balloons and their equipment, besides being expensive, require elaborate preparation for their flight [17] and even the best designed ones have not been able, till now, to penetrate the topmost layers of the ozonosphere. Further, they are inconvenient for day-to-day measurements. The most effective and practical method for studying the ozonosphere which has been developed in recent years, since its first application by Fabry and Buisson in 1921 [18], is the spectro-photometric method [19]. The essential principle of the method—as its name implies—consists in comparing the intensities of the direct or scattered sunlight in a region near the strong absorption bands of ozone, for different zenithal positions of the sun. Proper interpretation of the results yields the effective thickness, the height as also the distribution with height of the ozone layer.

(b) Sounding balloon method

The earliest of the ozone measurements by sounding balloon is that of E. and V. H. Regener [7, 20]. On June 26, July 7 and 31 in 1934,

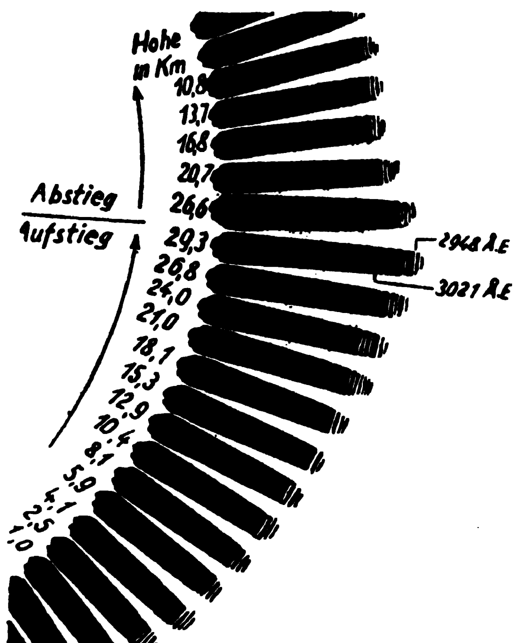


FIG. 4. Solar spectrum recorded in spectrograph carried by Regeners' balloon. As the balloon rose, more and more of the absorbing mass of ozone was left below and the spectrum extended towards the ultraviolet.

the Regeners sent up at Stuttgart, Germany, registering balloons carrying automatic quartz spectrograph, which reached up to heights 21, 20 and 31 km. respectively. The photographic plate registering the spectra moved every few minutes. Two barographs and a thermograph also recorded pressure and temperature on the plate.

Fig. 4 shows the photographs of solar spectrum as obtained on July 31, when the balloon attained the highest altitude. The extension of the spectrum towards the ultraviolet with the rise of the balloon is clearly seen.

The amount of absorption at any height and hence the thickness of the ozone layer above that height is obtained from the steepness of the darkening of the plate towards the ultraviolet and the known absorption coefficients.

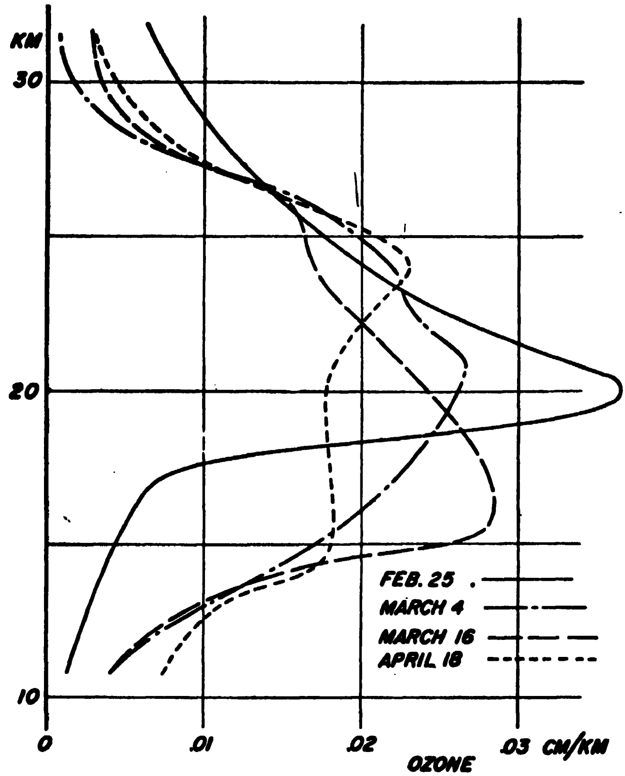


FIG. 5. Ozone distribution with height as calculated from steepness of blackening in the ultraviolet of the solar spectrum photographed in balloon flights made at New Mexico, U.S.A., in 1950. The abscissa indicates the thickness, reduced to S.T.P., of the ozone as exists in a layer one kilometre thick at a given altitude. (After Regener.)

It will be noticed that at 29·3 km.—the highest point where the spectrum was photographed—the plate is blackened up to 2875Å. The limit was reached with an exposure of 10 minutes only with spectrograph of small light gathering power carried in the balloon. On the other hand, with an

apparatus of large light gathering power and with several hours of exposure at Arosa, Götz [21] was able to extend the solar spectrum up to 2863Å, only a few Ångströms lower than the limit reached by the Regeners. It should, however, be mentioned that when the ozone content is small the spectrum can be extended towards the ozone absorption region with comparatively small exposure. Chiplonkar [21 *a*], for instance, working at Bombay was able to record the solar spectrum up to 2874Å, with an exposure of only 10 to 15 minutes using a simple Dobson spectrograph.

Ozone measurements from solar spectral photographs obtained in balloon flights have also been made at New Mexico, U.S.A., by V. H. Regener [17]. The balloons went well above the ozone maximum and the highest altitude reached was 32 km. Ozone curves as evaluated from the results of four flights (1950) are depicted in Fig. 5. The trends of the curves generally follow those obtained previously and show the marked day to day variation of the distribution and total ozone content. There were strong winds in the stratosphere on the days of the balloon flights. But no immediate relationship between the ozone distribution and the wind was apparent.

Extension of the solar spectrum beyond the ozone absorption limit as recorded in the V2 rocket flight and the ozone distribution curve deduced therefrom will be discussed in Chap. XII. It may be mentioned here that a dip at a height of about 21 km. has been observed in the ozone-distribution curve from the rocket flight. Such double-peak distribution, associated with abnormally high total ozone content, has also been recorded in observation by the 'Umkehr' method (*vide infra*).

(*c*) Spectro-photometric methods

(i) *The earlier method of Fabry and Buisson.*—Fabry and Buisson [18] were the first to determine the equivalent thickness of the ozone layer.

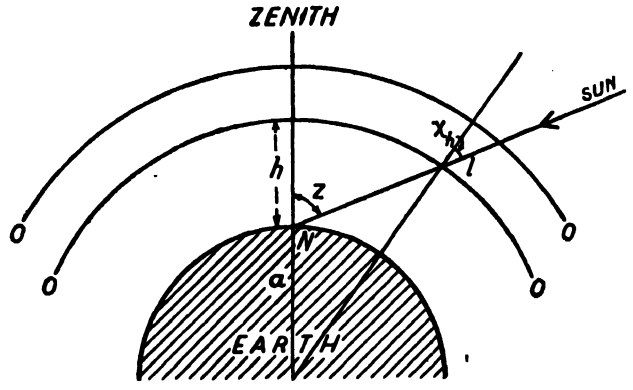


FIG. 6. Illustrating the principle of determining the equivalent thickness (x) of the ozone layer after Fabry and Buisson.

The principle of their method will be evident from Fig. 6. OO represents the ozone layer at height h above the surface of the earth. The meanings

of the other symbols are evident. Let the optical thickness of the layer be x for vertical incidence ($Z = 0$) and l for incidence at an angle Z. The following relations hold as long as Z is not very large:

$$l = x \sec \chi_h \quad \text{and} \quad \sin \chi_h = \frac{\sin Z}{1+h/a}.$$

It is evident that as the zenith angle Z increases, the solar rays reaching the observer have to traverse greater and greater thickness through the ozone layer. Now, if I be the intensity of a wavelength of direct solar light measured at the surface of the earth and I_0 that before its incidence on ozone layer, then

$$\begin{aligned}\log I &= \log I_0 - \alpha l \\ &= \log I_0 - \alpha x \sec \chi_h \qquad \ldots \quad \ldots \quad \ldots \quad (1)\end{aligned}$$

where α is the absorption coefficient of ozone for the wavelength under consideration. In principle therefore, x—the equivalent thickness—can be determined from two observations on the intensity of a particular wavelength measured at two zenithal positions of the sun.

In the actual experiment the solar spectrum is photographed for various zenithal positions of the sun and intensities of a number of narrow wavelength ranges near the ultraviolet end of the spectrum (where the ozone absorption begins) are measured. Ratio of the intensities of a wave band for any two zenithal positions of the sun gives the data necessary for calculating the ozone content. For greater accuracy a number of such comparisons are made. It should be noted that in conducting one set of observations throughout a whole day it is assumed that the ozone content does not vary from hour to hour.

(ii) *Improvements by Dobson and Harrison.*—In the above simple method it has been tacitly assumed that the only agency for reducing the intensity of solar rays traversing the atmosphere is ozone absorption. There are, however, other factors which cause reduction of the intensity of the rays. The solar rays in their passage through the atmosphere suffer Rayleigh scattering, i.e., scattering by the air molecules. Nearer the surface of the earth there is also a large amount of loss due to scattering by large particles of dust and water. It is easy to see that both these scatterings depend upon the position of the sun. Fabry and Buisson in their experiments took account of loss by molecular scattering but did not do so for scattering by large particles. In the method of computing the equivalent amount of ozone devised by Dobson and Harrison [22] to be described presently, the effect of scattering by large particles is automatically eliminated. The principle of the method is as follows: For any wavelength we have from Eq. (1),

$$\log I_0 - \log I = \alpha x \sec \chi_h,$$

if only ozone absorption is present. When losses due to the scattering processes are considered the above relation becomes

$$\log I_0 - \log I = (\beta+\delta) \sec Z + \alpha x \sec \chi_h, \qquad \ldots \quad (2)$$

where β and δ are scattering coefficients of the whole atmosphere ($Z = 0$) for Rayleigh scattering and for scattering by large particles respectively.

For a second wavelength,

$$\log I_0' - \log I' = (\beta' + \delta) \sec Z + \alpha' x \sec \chi_h. \quad .. \quad (3)$$

δ is assumed to be same for wavelengths not much separated. From (2) and (3) we get finally

$$\log I/I' = \log I_0/I_0' - (\alpha - \alpha')x \sec \chi_h - (\beta - \beta') \sec Z. \quad .. \quad (4)$$

Thus, relation (4) does not involve δ and the value of x may be determined if we know the values of α, α', β, β', $\log I/I'$, $\log I_0/I_0'$ and χ_h. Of these α and α' may be obtained from laboratory experiments. The other quantities may be evaluated if a series of direct sun observations for $\log I/I'$ is made on a day of uniform atmospheric conditions. It is of course assumed that during the period of observation x and h remain constant. In Fig. 7, curve (a) is a typical plot of a series of observations showing the variation

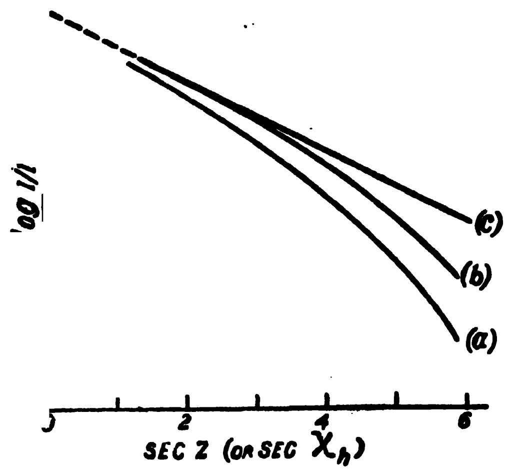


FIG. 7. Illustrating how x and h may be obtained from a plot of $\log I/I'$ against sec Z (curve (a)) I is the intensity of light strongly absorbed (λ 3110) and I' that weakly absorbed (λ 3290). (After Götz, Meetham and Dobson.)

of $\log I/I'$ with sec Z. The process of obtaining x and h from this curve is as follows [6]:

Curve (b) is first derived from curve (a) by adding to each ordinate the corresponding $(\beta - \beta') \sec Z$. (The values of β and β' are obtained theoretically from Rayleigh's formula for molecular scattering.) Curve (b) thus represents the variation of $\log I_0/I_0' - (\alpha - \alpha')\, x \sec \chi_h$ with sec Z.

The abscissa sec Z is now replaced by sec χ_h and curve (c) is drawn. Now, this curve (c) ought to be a straight line if the value of h assumed in sec χ_h is its true value. In actual practice therefore, a number of curves such as (c) is drawn for different assumed values of h and the value corresponding to the particular curve (c) which is a straight line, is taken as the true value of h—the height of the ozone layer.

The finally selected c-curve, produced to meet the axis of $\log I/I'$, gives the point which corresponds to the value of $\log I_0/I_0'$. x is finally determined from the slope of the line (c) which is $(\alpha - \alpha')x$. In practice, the

values of log I_0/I_0' and h are determined for a number of days in uniform atmospheric condition.

(iii) *Use of photo-electric cell—the photo-electric apparatus* (*Dobson*).—A notable advance in the method of comparison of spectral intensities for atmospheric ozone measurement was made by Dobson [19] in 1931 by the introduction of photo-electric cells in place of photographic plates. The photographic methods, though they have some advantages, have disadvantages also for daily routine measurements. They are therefore now discarded in favour of the photo-electric method. This method besides being very quick (only about 5 minutes sufficing for a complete set of readings) can be used either for the direct sunlight or for light scattered from the zenith sky and may even be employed when the sky is overcast with clouds. We shall see presently that the development of the photo-electric apparatus by Dobson has made possible the determination of ozone distribution with height by observation on the scattered light from the zenith sky.

The general principle of the photo-electric apparatus is as follows: A double quartz spectroscope isolates two narrow bands in the ultraviolet so chosen that the longer wavelength is very little affected by ozone absorption while the shorter one is greatly absorbed. Measurement of the ratio of intensities of these two bands allows one to calculate the amount of ozone through which the sunlight has passed. In order to determine this ratio, the radiations corresponding to the two wavelengths (3110A and 3290Å) are made to fall alternately on a photo-electric cell, the alternations being caused by a shutter rotating at a speed of 10 to 20 times per second. The less absorbed wavelength can be reduced in intensity by an adjustable optical wedge. The fluctuating output of the cell is amplified by a four-stage low frequency valve-amplifier. If by adjustment of the wedge, the intensities of the two wavelengths falling alternately on the cell be equalized, then a steady current will flow into the input of the amplifier and no amplification will result. If therefore the wedge is calibrated beforehand, the ratio of intensities of the two wavelengths may be obtained from its setting when there is no amplification by the amplifier. The arrangements of the optical parts, the shutter and the photo-cell are shown in Fig. 8.

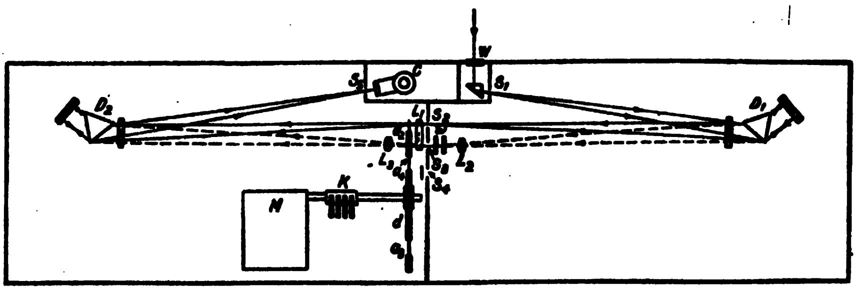


FIG. 8. Illustrating the principle of Dobson's photoelectric apparatus for estimating the ozone content from observations on the scattered light from the zenith sky.

Light enters the instrument through the window W and after passing through the slit S_1 falls on the first dispersing system D_1. The slits S_2, S_3 and S_4 isolate three narrow bands of wavelengths 3110, 3290 and 4435Å respectively (the third one, passing through S_4, is for measuring the transparency of the atmosphere for wavelengths not affected by ozone absorption). The rays of wavelengths to be compared, on emergence from slits S_2 and S_3, fall on the second dispersing system D_2 which is similar to D_1 and recombine on slit S_5. The lenses L_1, L_2 and L_3 ensure that the rays are properly focussed and pass through the slits accurately. For reducing the intensity of the longer wavelength by a known amount, two optical wedges ω of neutral tinted gelatine are placed behind the slit S_3. d is the sector wheel with apertures a_3 and a_4 which revolves close to the slit system and admits light alternately from S_2 and S_3. K is the commutator and M the driving motor for rotating the sector wheel. The commutator is mounted on the shaft carrying the sector and its purpose is to reverse the direction of the current from the amplifier output at right times and thus convert the alternating current into a pulsating unidirectional current which is read on a D.C. microammeter.

The reason for using a double spectrograph is this. Since the intensity of daylight in the region 3110Å is very small compared to that at 3290Å, the light of latter wavelength scattered by the lens and prism surfaces would fall in appreciable proportion in the photo-cell. With a double spectrograph, however, the scattered light is dispersed again and only a negligible amount falls on the cell.

(iv) *Determination of vertical distribution of ozone—Umkehr Effect.*— We discussed in the previous section the methods of determining the equivalent thickness of the ozone layer and its average height. It is now natural to enquire how the ozone is distributed vertically in the atmosphere. A method of doing this was suggested by Götz [23] in 1929. He observed that the light scattered from the zenith sky is relatively richer in short wavelengths when the zenith distance of the sun is about 85° than when it is smaller. He showed that if observations are made on the light scattered from the zenith sky after the photo-electric method of Dobson, the curve depicting the variation of $-\log I/I'$ with Z^4 (Z—zenith distance) is of the form shown in Fig. 9. It will be noticed that the ratio of intensities (I/I') of the two wavelengths first decreases with the increase of zenith distance, attains a minimum when the zenith distance is about 86° and then increases. In other words, the scattered light from the blue zenith sky is comparatively richer in ultraviolet when the sun is setting than when it is higher up. The effect is called by Götz '*Umkehr Effect*' and the curve the '*Umkehr Curve*'. Before discussing the actual method of computing the vertical distribution of ozone with the help of Umkehr curve, it would be useful to explain how this inversion takes place.

The scattered light from the zenith sky differs in one important respect from that transmitted directly from the sun. In the latter the length of path through the ozone layer always increases with the zenith distance.

Thus the ratio of intensities of two lines, 3110Å and 3290Å (the former greatly and the latter little absorbed) always decreases with the increase

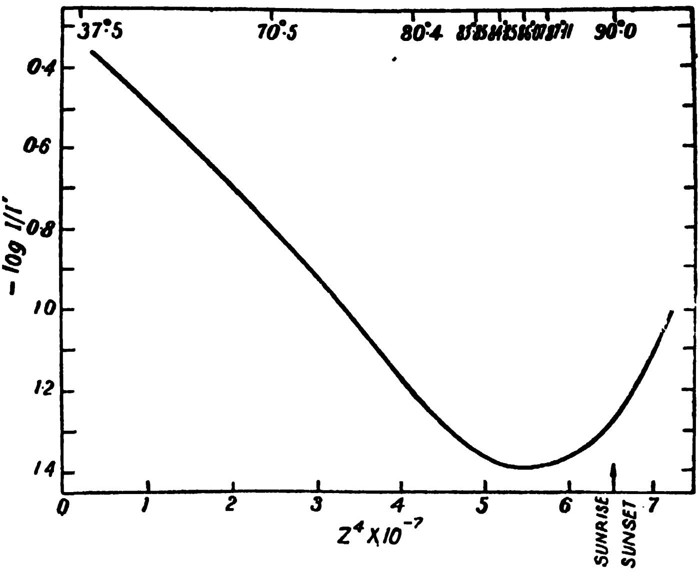


FIG. 9. Illustrating the Umkehr Effect. Note how the intensity ratio I/I' (compare Fig. 7) at first decreases with the increasing zenith distance of the sun and that immediately before sunset (or sunrise) it begins to increase. In the upper part of the figure the zenith angles of the sun corresponding to the values of $Z^4 \times 10^{-7}$ are given. (After Götz, Meetham and Dobson.)

of zenith distance. In the case of the light from zenith sky there is scattering at all heights from the direct solar beam and the amount of light from any particular level reaching the ground will depend not only upon the length of path through the ozone layer but also on the amount of air available at that height for scattering. Thus, consider a thin layer at A in the region of ozone (Fig. 10). The amount of light from the layer will depend, firstly on the density of air at A and, secondly, on the absorption by ozone before and after scattering. Now consider another thin layer at B above A. Here the effect of the first will be to reduce and that of the second to increase the amount of light from B. The latter because the longer oblique path before scattering through ozone is replaced by shorter vertical path (after scattering). The reverse is the case when we consider a layer below A. Thus when the contribution from each layer is considered separately, the total amount of scattered light reaching the instrument is found to depend on two factors varying in opposite senses. One of these causes an increase and the other a decrease in the amount of light received as one goes higher up. Now, since at the surface of the earth we receive the integrated light from all the layers, it is obvious that due to the two counteracting effects most of the scattered light

reaching the instrument will be confined to a limited space. The height of this region may be called the *effective height* of scattering. It is evident that the effective height of scattering is greater when the

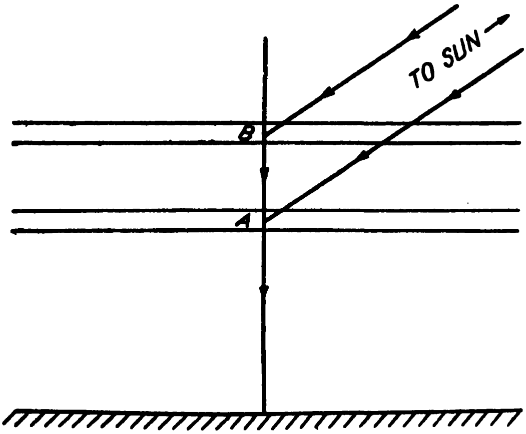


FIG. 10. Illustrating how the intensity of the light scattered from the zenith sky depends upon the position of the scattering layer within the ozone region.

zenith distance of the sun is large and also for the wavelengths which are strongly absorbed by ozone. Consider now the two wavelengths—3110Å and 3290Å—the former strongly and the latter weakly absorbed by ozone. The effective height for 3110Å is always higher than that for 3290Å. Now, near about sunset when the sun is a few degrees above the horizon the effective height for 3110Å will be above the whole or at least the greater part of the ozone layer, while 3290Å will be scattered much lower down. In such case, most of the absorption of 3110Å by ozone takes place only in the short vertical path. Since this absorption path practically remains constant as the sun goes down the intensity of 3110Å decreases much more slowly than that of 3290Å. As a result of this the ratio of the intensities of these two wavelengths at first decreases with the increase of zenith distance of the sun; then, after attaining a minimum, increases again for larger values of the zenith distance. It should, however, be noted that it is only the ratio of intensities of the two wavelengths which increases and not the absolute intensity. This latter always decreases with the sinking of the sun.

Observations on days of uniform atmospheric conditions show that the shape of the Umkehr curve and therefore the vertical distribution of ozone depends mainly on the total ozone content.

It is to be noted that there is no strict mathematical method of finding the vertical distribution of ozone from a given Umkehr curve. Two different approximate methods have been developed, one at Arosa and the other at Oxford [6]. They give similar results and the principle of the method used at Arosa will now be briefly described.

In this method the atmosphere is conveniently divided into five sections:—

(*a*) Region between 2–5 km.
(*b*) „ „ 5–20 km.
(*c*) „ „ 20–35 km.
(*d*) „ „ 35–50 km.
(*e*) „ above 50 km.

It is assumed that in section (*e*) the amount of ozone is negligible. In accordance with measurements at the ground level the amount in section (*a*) is taken as 3 per cent (u) of the total amount. In section (*b*) the amount is $x-(x_1+x_2+u)$, where x_1 and x_2 are the amounts in sections (*d*) and (*c*) respectively and x—the total amount obtained from direct sun observations. It is further assumed that the ozone in each section is uniformly distributed.

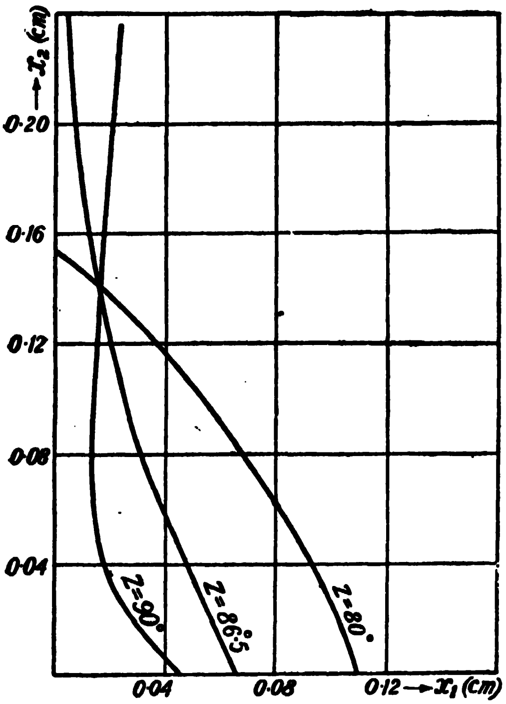


FIG. 11. Illustrating how the ozone contents in the regions 50–35 km. (x_1) and 35–20 km. (x_2) can be evaluated by the graphical method of solution. (After Götz, Meetham and Dobson.)

Thus there are two unknowns—x_1 and x_2—to be evaluated. For this purpose the values of log I/I' for two different values of Z are calculated as follows:

The amount of light dI reaching the instrument which was scattered vertically downwards at any height is given by

$$dI = 10^{-\beta L} \,.\, 10^{-\alpha l} \,.\, b \times \text{const.}$$

Here the first factor on the right is proportional to the intensity of the light reaching the height in question by scattering; the second factor takes

into account the absorption by ozone of this light in reaching the instrument by being scattered vertically downwards. The symbols are:

β—scattering coefficient of the whole atmosphere for 760 mm. pressure,

L—total mass of air traversed by the radiation
$= 1 + b/760\,[f(Z)-1]$,

b—pressure in mm. at the height of scattering,

$f(Z)$—Bemporad's function which tends to sec Z as $Z \to 0$,

α—absorption coefficient of ozone,

l—length of path through ozone.

Hence, the intensity I measured at the ground (of 3110Å say) is given by

$$I = \Sigma 10^{-\beta L} . 10^{-\alpha l} . b \times \text{const.}$$

The above summation is carried out in steps of one kilometre from 2 to 65 km. It is seen that l is a function which contains the two unknowns x_1 and x_2, and also x and u, and may be calculated trigonometrically. The values of log I/I' for any two values of Z may then be deduced. From the observed Umkehr curve also two values of log I/I' corresponding to the above two values of Z may be obtained. Two simultaneous equations are thus obtained which may be solved graphically as shown in Fig. 11 where x_2 is plotted against x_1 for each value of Z. The point of intersection of the two curves corresponding to two values of Z gives the numerical solution of the two simultaneous equations. As a check, a third curve for a different value of Z is drawn.

Evaluations of x_1 and x_2 give the 'block' distributions shown in Fig. 12. A smooth curve is then drawn keeping the amount of ozone in each section equal to that in the corresponding block. This distribution may be compared with that obtained experimentally by the Regeners [7, 17, 20] (see Fig. 5).

It will be seen that the average height of the ozone layer lies between 22 and 25 km. This differs from observations made by previous workers which gave a value of about 40 km. It is further seen that ozone is not confined in a thin layer but is spread between the heights 5 to 45 km.

It is to be noted that the most probable distribution as obtained by the Umkehr method and as depicted by the smooth curves in Fig. 12 is not really so accurate as it appears to be. The height of the centre of gravity, however, is more accurate and can be relied to within a kilometre.

The fact that most of the ozone exists between 10 km. and 35 km. may seem to contradict the hypothesis that the high temperature of the middle atmosphere in the region of 45 km. (as evidenced by the phenomena of abnormal propagation of sound and of meteoric flashes, Chapters IIIA and IIIB) is due to absorption of solar radiation by ozone. It will, however, be shown in Sec. 5 that owing to the very large coefficient of absorption the heating effect of the absorbed radiation is confined to the uppermost layer of the ozone distribution.

(v) *Other methods.*—The average height of the ozone layer can also be determined by a simple method first worked out by Strong [23*a*]. It has been

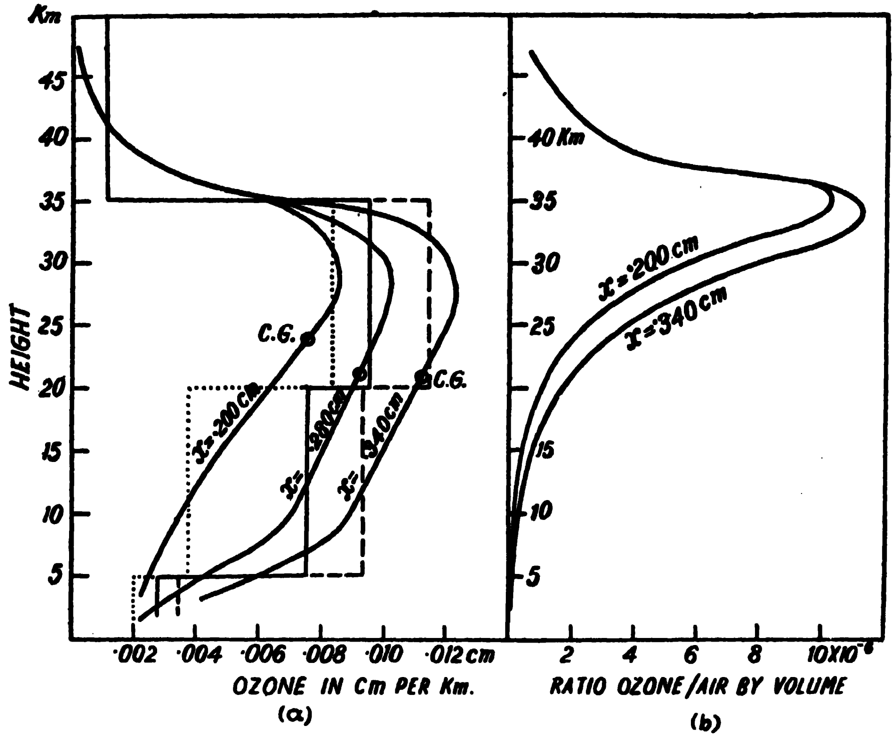


FIG. 12. Variation of ozone content with height as determined by Umkehr Effect. The curves are for Arosa (Switzerland). In (*a*), it is seen that the greatest absolute amount of ozone lies between 25 and 30 km. The ratio of ozone to air by volume is, however, not greatest at this height (see (*b*)). This is because of the decrease of atmospheric density with height. (After Götz, Meetham and Dobson.)

shown by Strong that the absorption by ozone in the infra-red 9·7 μ band, varies, for a constant amount of ozone, as the fourth root of the total pressure under which the ozone exists. The atmospheric absorption of the solar infra-red radiation in this band is measured, and at the same time the total ozone content by absorption measurement at the long wavelength edge of the Hartley band, as already described, is determined. From these two determinations the average pressure under which the ozone existed can be found and hence also the average height. The method does not give the distribution with height but has the advantage that the determination is made with the help of one set of readings only. The change in the total ozone content, as also the change in the average height, can thus be followed through the daylight period. It is to be mentioned that the method employs direct sunlight and, is thus, inapplicable on cloudy days.

A novel method of determining the ozone distribution with height from observations of the earth's shadow on the moon during a lunar eclipse is described in Section 9.

4. VERTICAL DISTRIBUTION—THEORETICAL METHODS

(a) Theories of formation and destruction of atmospheric ozone

Formation of ozone.—The primary reaction in ozone formation is the dissociation of O_2. The oxygen atoms formed combine with O_2 to produce O_3 thus [24],

$$O_2 + h\nu \rightarrow O + O$$

$$O + O_2 + M \rightarrow O_3 + M$$

where M is the third body which absorbs the extra energy and momentum [see Chapter X, Sec. 4(*d*)].

The collision of an excited O_2 molecule with O_2 may also produce ozone thus,

$$O_2^* + O_2 \rightarrow (O_2 + O) + O$$

$$= O_3 + O$$

$$O + O_2 + M \rightarrow O_3 + M.$$

The first reaction means that simultaneously with the dissociation, one of the oxygen atoms produced combines with the reacting O_2 to produce O_3. The other oxygen atom then combines with O_2 by three-body collision process.

Now, O_2 has the following absorptions in the ultraviolet which may produce O_3 by the above reactions:

(1) Herzberg bands.—These are weak absorption bands in the ultraviolet converging near 2400Å. Warburg [24] was able to establish the formation of ozone by light of wavelength 2530Å. The radiation of this wavelength is unable to dissociate O_2 and the production of O_3 is accounted for by the following reactions:

$$O_2 + h\nu \rightarrow O_2(^3\Sigma_u^+)$$

$$O_2(^3\Sigma_u^+) + O_2 \rightarrow O_3 + O$$

$$O + O_2 + M \rightarrow O_3 + M. \quad \ldots \quad \ldots \quad \ldots \quad \text{(I)}$$

(2) Absorption in the Runge-Schumann *continuum*.—This absorption begins at 1760Å and goes up to 1250Å [see Appendix 2(*f*)]. The absorption dissociates O_2 thus,

$$O_2 + h\nu \rightarrow O(^3P) + O(^1D). \quad \ldots \quad \ldots \quad \ldots \quad \text{(II)}$$

The absorption is very strong having a maximum at 1450Å where, according to Ladenburg and Voorhis [25] a thickness of only 0·0014 cm. at S.T.P. reduces the intensity to half. This is stronger than the maximum absorption of ozone in the region 2550Å.

(3) Runge-Schumann *absorption bands*.—These absorption bands begin at about 1925Å and converge to 1760Å. The absorption excites the oxygen molecule thus,

$$O_2 + h\nu \rightarrow O_2^*(\overline{O^3P + O^1D}\, B\, {}^3\Sigma). \quad \ldots \quad \ldots \quad \text{(III)}$$

Further, according to Flory [26] this absorption may lead to predissociation so that oxygen atoms may also be produced. The absorption coefficient in

this region is not accurately known, though it is found to be much smaller than that of the Runge-Schumann continuum. According to Kreusler [27] a thickness of 20 cm. of O_2 at S.T.P. will reduce λ1860 to two-thirds of its intensity. The coefficient is smaller for longer wavelengths. The excited molecule or the O atom produced by this absorption can react with normal O_2 to produce ozone.

We now discuss which of the three absorptions listed above can produce the atmospheric ozone layer [28]. Clearly, reaction (I) is not helpful as the range of wavelength lies within the strong ozone absorption bands. The Runge-Schumann continuum (reaction (II)) is also not concerned with the production of the ozone layer. The radiation in this range is entirely used up above 80 km. in dissociating O_2 [Chapter V]. In this high region there is also little chance of three-body collision on account of low density. According to calculation by Chapman [29], the minimum pressure for three-body collision process is 10^{-2} mm. of mercury and this is attained at a height of 80 km. Oxygen atoms above this level will therefore produce little ozone even if oxygen molecules are present. Conditions are, however, different with reaction (III), namely, the Runge-Schumann absorption bands. Since the absorption coefficient is small, radiations in this wavelength range will penetrate deep into the atmosphere. As the equivalent thickness of O_2 present above 45 km. is about 20 cm., λ1860 will reach in sufficient intensity to produce excited O_2 at the level. The pressure here is also sufficiently high to ensure three-body collision. Further, at 35 km. height the equivalent thickness of the overlying oxygen is 840 cm. and the intensity of λ1930 is reduced to one-tenth by this absorption. These considerations show that the presence of ozone in the middle atmosphere can be accounted for by the absorption of solar ultraviolet radiation in the region of Runge-Schumann bands. The region in which ozone is formed according to the above estimate is higher than the region of maximum concentration of ozone. According to some authorities ozone may be produced in the high region as estimated above and carried down to lower levels by convection.

Destruction of ozone.—Ozone is destroyed by photo-dissociation (photolysis) and also by collision with atomic oxygen. Thus, for photolysis

$$O_3+h\nu \rightarrow O_2+O^*.$$

It has been observed in the laboratory that even red light λ6200 is able to decompose ozone [30]. But so far as the ozonosphere is concerned, the spectral region mainly responsible for the destruction of ozone is that in the range of the Hartley bands. The oxygen atom which is produced reacts with ozone thus,

$$O_3+O \rightarrow 2O_2^*.$$

Mention should also be made of photolysis of ozone in the presence of water vapour. Water vapour acts as catalyst and, if present in ozonized oxygen, increases the quantum yield (i.e. the number of oxygen molecules decomposed per quantum absorbed by ozone). This was first observed by Warburg [28(*a*)] and has since been confirmed by experiments of Forbes

and Heidt [28(*b*)]. These authors investigated the photolysis of ozone in the presence of saturated water vapour by radiations in the wavelength range of the Hartley bands. The partial pressure of ozone was varied from 10 mm. to one atmosphere and the temperature from 2°C. to 20°C. It was found that the quantum yield is proportional to the concentration of water molecules.

(For a discussion of the photolysis of ozone the reader may refer to [8].)

(*b*) Theoretical calculation of the vertical distribution of atmospheric ozone

It is possible to calculate theoretically the vertical distribution of ozone from a knowledge of the photochemical reactions and the absorption coefficients in the spectral regions which are responsible for ozone formation and decomposition. Such calculations have been made by Chapman [29], by Mecke [31] and by Wulf and Deming [32]. The calculations of the first two authors were made when the maximum concentration of ozone was believed to be at 40 to 50 km. These calculations have not been revised. We describe below briefly the analysis of Mecke and of Wulf and Deming.

(i) *Method of Mecke.*—We have already described the nature of the photochemical reactions which produce and destroy ozone. In the condition of equilibrium we have:

$$O_2+O_2 \rightleftarrows O_3+O$$

and

$$O_2+O \rightleftarrows O_3.$$

As already explained the origin of these reactions is the photochemical action of the solar ultraviolet radiation. (In the above, the intermediate reactions have been omitted and only the end products are shown.) Inserting the reaction constants we may write the above processes as follows:

$$O_2+O_2 \underset{k_3}{\overset{p_{O_2}a_2}{\rightleftarrows}} O_3+O \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \text{(I)}$$

$$O_2+O \underset{a_3}{\overset{k_2}{\rightleftarrows}} O_3 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \text{(II)}$$

Combining reactions (I) and (II),

$$3O_2+O \underset{k_3a_3}{\overset{p_{O_2}a_2k_2}{\rightleftarrows}} 2O_3+O \quad \ldots \quad \ldots \quad \ldots \quad \text{(III)}$$

where

p_{O_2}—the quantum efficiency of the process,

$a_2=\epsilon_2 I_2$ and $a_3=\epsilon_3 I_3$, in which I_2 and I_3 are the intensities of radiations which produce and decompose ozone respectively and ϵ_2, ϵ_3 are their absorption coefficients per Mol./c.c.,

k_2, k_3—reaction constants of collisional processes with O which produce and decompose ozone respectively.

If τ be the mean lifetime of the excited oxygen molecule, whether it comes to the ground state by radiation or by collision, then we have

$$p_{O_2} = \frac{(\tau k_1)\,[O_2]}{1+(\tau k_1)\,[O_2]},$$

where k_1 is the reaction constant for the process

$$O_2^* + O_2 \longrightarrow O_3 + O.$$

If the pressure be small $(\tau k_1)\,[O_2] \ll 1$, so that

$$p_{O_2} \fallingdotseq \tau k_1\,[O_2].$$

Substituting for p_{O_2} we get from reaction (III)

$$[O_3]^2/[O_2]^3 = \frac{\tau k_1 k_2 a_2}{k_3 a_3} = K \text{ (say)}. \quad \ldots \quad \ldots \quad (5)$$

The distribution of ozone may now be determined in terms of the maximum ozone concentration. Since reactions (II) and (III) necessitate three-body collision, the ratio k_2/k_3 depends on pressure and hence from Eq. (5) K may be put as

$$K = K'p\, I_2/I_3$$

where K' is a constant, and p the atmospheric pressure.

Now I_2 and I_3 at any height h above the earth's surface may be put as

$$I_2 = I_{0,2} \exp\left(-\int_h^\infty \epsilon_2[O_2]\,dh\right)$$

and

$$I_3 = I_{0,3} \exp\left(-\int_h^\infty \epsilon_3[O_3]\,dh\right),$$

so that

$$K = C_1 p\,.\exp\left(-\int_h^\infty \left\{\epsilon_2[O_2]-\epsilon_3[O_3]\right\}dh\right), \quad \ldots \quad (6)$$

where C_1 is another constant. Assuming that the density variation of O_2 follows the barometric law,

$$[O_2] = C_2 p, \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (7)$$

where C_2 is a constant. From (5), (6) and (7),

$$[O_3]^2 = C_1 C_2^3 p^4 \exp\left(-\int_h^\infty \left\{\epsilon_2[O_2]-\epsilon_3[O_3]\right\} dh\right)$$

$$= Cp_0^4 \left(e^{-\frac{h}{H}}\right)^4 \exp\left(-\int^\infty \left\{\epsilon_2[O_2]-\epsilon_3[O_3]\right\} dh\right) \quad \ldots \quad (8)$$

since $p = p_0 e^{-\frac{h}{H}}$ where H = scale height of the homogeneous atmosphere. Differentiating (8),

$$\frac{1}{[O_3]}\frac{d[O_3]}{dh} = -\frac{2}{H} + \frac{1}{2}\epsilon_2[O_2]\left\{1 - \frac{\epsilon_3[O_3]}{\epsilon_2[O_2]}\right\}. \quad \ldots \quad \ldots \quad (9)$$

Now $[O_3]/[O_2]$ is not greater than 10^{-3} and the quantity within brackets on the right hand side may be put approximately equal to unity so long as ϵ_3/ϵ_2 is not 10^3. Therefore

$$\frac{1}{[O_3]}\frac{d[O_3]}{dh} = -\frac{2}{H}+\frac{1}{2}\epsilon_2[O_2]. \quad .. \quad .. \quad .. \quad (9.1)$$

The O_2-concentration at the height where the concentration of O_3 is maximum is now obtained from

$$\frac{2}{H} = \frac{1}{2}\epsilon_2[O_2]_m,$$

or,

$$[O_2]_m = \frac{4}{\epsilon_2 H}. \quad .. \quad .. \quad .. \quad .. \quad (10)$$

Integrating (9.1), and with the help of (10), we get

$$[O_3] = [O_3]_m\left(\frac{p}{p_m}.e^{1-(p/p_m)}\right)^2, \quad .. \quad .. \quad .. \quad (11)$$

where p_m is the pressure at the level of maximum ozone concentration ($[O_3]_m$).

From (11) the total ozone content may be easily obtained by integration. Thus,

$$[O_3]_{total} = 1{\cdot}85[O_3]_m p_m.$$

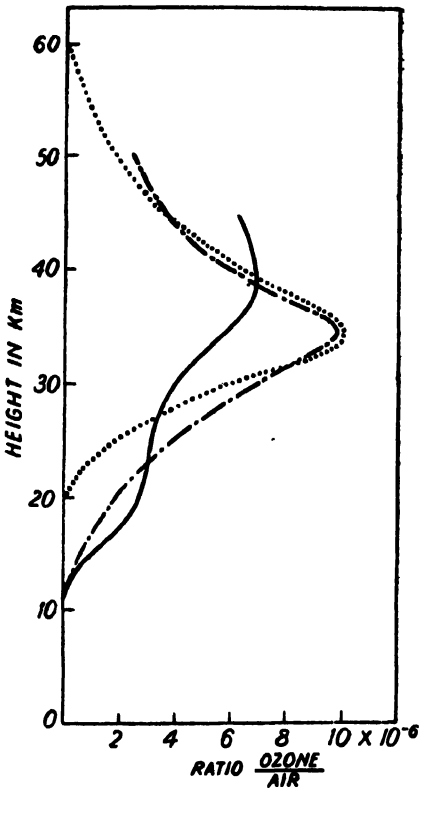


FIG. 13. Comparison of the vertical distribution of ozone after Mecke's calculation (dotted curve) with distributions as obtained by the Umkehr method at Tromsö (full line curve) and at Arosa (.—.—.—.). (After Penndorf.)

In Fig. 13 ozone distributions as actually obtained by the Umkehr method at Tromsö and at Arosa are compared with the distribution curves obtained theoretically after Mecke. It will be seen that for Arosa the two curves are in excellent agreement. For Tromsö, however, the observed curve departs greatly from the calculated one.

In the above derivation of ozone distribution several simplifying assumptions have been made and some disturbing factors are neglected. The absorption coefficients of the radiations vary over a wide range in the spectral ranges which lead to the photochemical reactions. It is therefore not justifiable to take their mean values assuming them to be constant over the spectral ranges considered. Further, the density distribution (Eq. (7)) is independent of temperature. This is clearly wrong in view of the fact that there is a rise of temperature with height in the ozone layer. Finally, the effect of possible diffusion is not taken into account. Notwithstanding these assumptions the simple and direct method of attack gives an insight into the physical processes involved.

(ii) *Method of Wulf and Deming.*—The assumptions on which this method of calculating the distribution of atmospheric ozone is based are the following:—

(*i*) Ozone exists in an undisturbed photochemically steady state under the influence of standard solar radiation (6000°K).

(*ii*) The atmosphere is in diffusive equilibrium above 20 km. and is at a constant temperature of 219°K.

(*iii*) Solar ultraviolet radiations of $\lambda < 2400$Å decompose O_2-molecules into O-atoms; the reactivity of O-atoms in the processes of formation and decomposition of ozone is approximately the same whether they are normal or excited.

(*iv*) The reactions which determine the steady states of atomic oxygen and of ozone concentration in the middle atmosphere are:

$$O_2 + h\nu \longrightarrow O + O \quad \ldots \quad \ldots \quad \ldots \quad \text{(I)}$$

$$O_2 + O + M \longrightarrow O_3 + M \quad \ldots \quad \ldots \quad \ldots \quad \text{(II)}$$

$$O_3 + h\nu \longrightarrow O_2 + O \quad \ldots \quad \ldots \quad \ldots \quad \text{(III)}$$

$$O_3 + O \longrightarrow O_2 + O_2 \quad \ldots \quad \ldots \quad \ldots \quad \text{(IV)}$$

where M is the third body carrying away extra energy and momentum.

For the above processes the equation representing the maintenance of equilibrium concentration of oxygen atoms is:

$$K_f n_1 n_2 n_M + K_d n_1 n_3 = 2\alpha_2 n_2 q_2 + \alpha_3 n_3 q_3 = 2Q_2 + Q_3 \quad \ldots \quad (12)$$

where

K_f, K_d—specific rates of ozone forming reaction (II) and ozone decomposing reaction (IV) respectively,

n_1, n_2, n_3—numbers per cm.^{3} of O, O_2 and O_3 respectively,

n_M—total number of molecules per cm.^{3},

α_2, α_3—absorption coefficients of the radiations which decompose O_2 and O_3 respectively,

q_2, q_3—corresponding numbers of quanta absorbed per cm.³ at any height h,

$Q_2 = \alpha_2 n_2 q_2$,

$Q_3 = \alpha_3 n_3 q_3$.

From (12)

$$n_1 = (2Q_2+Q_3)/(K_f n_2 n_M + K_d n_3).$$

Again, the condition of steady state concentration of O_3-molecules may be written as

$$K_f n_1 n_2 n_M = K_d n_1 n_3 + \alpha_3 n_3 q_3 = K_d n_1 n_3 + Q_3;$$

whence

$$n_1 = Q_3/(K_f n_2 n_M - K_d n_3).$$

Equating the above two values of n_1

$$n_3 = \frac{K_f}{K_d} \cdot n_2 n_M \left(\frac{1}{1+Q_3/Q_2}\right) \quad \cdots \quad (13)$$

The above equation shows that the number of O_3-molecules at any height may be obtained in terms of a constant K_f/K_d, of the concentration of O_2-molecules, of the total concentration of molecules at this height and of the numbers of quanta absorbed per cm.³ by ozone and oxygen molecules at this height.

Now the value of the constant K_f/K_d—the ratio of the specific reaction rates can be obtained from results of laboratory experiments on ozone equilibrium and may be taken as 10^{-19} after the experimental results of Eucken and Patat [33]. This numerical constant can also be obtained from the known total amount of ozone and the form of the absorption coefficient curves (Fig. 14) for O_3 and O_2. (In view of the uncertainty of the experimental results in the vicinity of the region 2000Å, the two absorption curves have been joined in three different manners labelled as C, D and E. For the cases C and E, K_f/K_d is taken as 10^{-19}, whereas for the case D, a value of 10^{-18} has been used in order to keep the area under the distribution curve approximately equal to the known total ozone in the vertical path.)

Since both α's and q's are functions of the frequency of radiation, Q_2 $(= \alpha_2 n_2 q_2)$ and Q_3 $(= \alpha_3 n_3 q_3)$ are obtained by summing over the entire frequency interval using finite strips $\Delta\nu$ in width sufficiently small for the purpose. Therefore, with the help of the known absorption coefficient curve (Fig. 14) and the number of quanta in different frequency ranges, the values of Q_2 and Q_3 can be obtained for different heights. It is to be noted that since Q_3 contains n_3, Eq. (13) is evaluated by the method of successive approximation. Fig. 15 depicts the three distribution curves for the three different cases C, D and E. It will be seen that the distribution depends to a large extent on the form of the absorption coefficient curves for O_2 and O_3. It is further seen that the calculated distributions are a little too high to correspond with the observed distribution of atmospheric ozone. This may be due to the fact that the value of the constant (K_f/K_d) depends on temperature which varies greatly in the middle atmosphere. It may be noted that Chapman has made estimates of the values of K_f

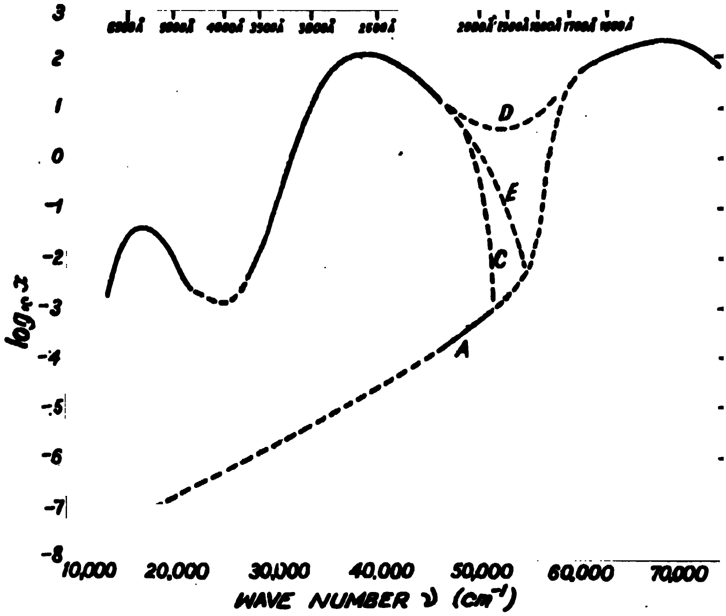


FIG. 14. Absorption curve of O_3 (left of D) and of O_2 (right of D). The region 60,000 cm.$^{-1}$ to 75,000 cm.$^{-1}$ is the Runge-Schumann continuum. Region below 60,000 cm.$^{-1}$ is the absorption due to Runge-Schumann bands which joins in some continuous manner with the region 47,000 cm.$^{-1}$ to 53,000 cm.$^{-1}$ (A). The region below (A) is extrapolated. The region 29,000 cm.$^{-1}$ to 42,000 cm.$^{-1}$ is the absorption due to ozone (the values are taken from measurements by Fabry and Buisson [9(a)], Läuchli [9(b)] and Ny-Tsi-Zé and Choong-Shin-Piaw [10]. The region 22,000 cm.$^{-1}$ to 29,000 cm.$^{-1}$ is extrapolated. (After Wulf and Deming.)

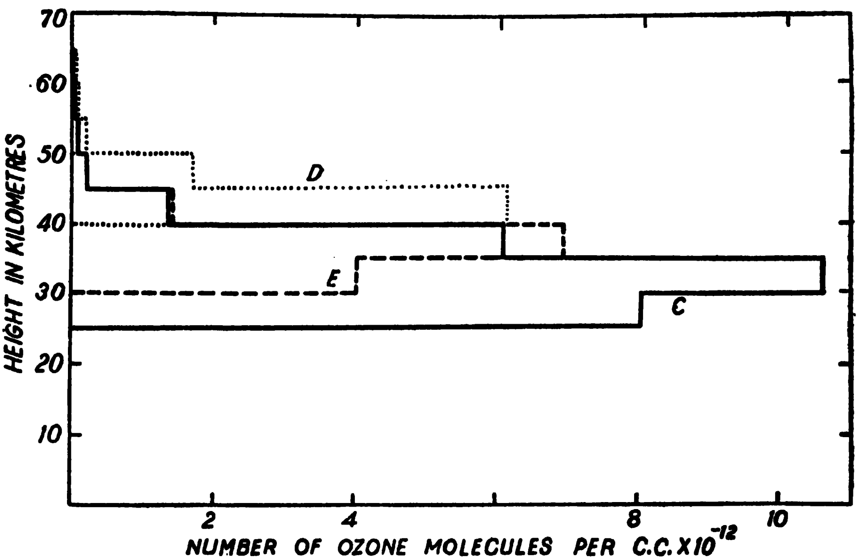


FIG. 15. Three theoretical distribution curves for ozone corresponding to the three assumed absorption curves C, D and E in Fig. 14. (After Wulf and Deming.)

and K_d (as also the coefficients for the production of O_2 by triple collision of two O atoms with third body) from purely solar and meteorological data. These estimates—$K_f = 2{\cdot}2 \times 10^{-37}$ and $K_d = 6 \times 10^{-20}$—are of reasonable orders of magnitude and their ratio agrees with the value adopted by Wulf and Deming in their calculation [8].

Wulf and Deming have, after considering the possible effects of any disturbance on the ozone-equilibrium, shown that in regions below the maximum of distribution, the time required for noticeable restoration of equilibrium is of the order of days. The ozone content of this region is thus subject to variations by such causes as atmospheric circulation. They are of opinion that two different views might be taken regarding the distribution of ozone in the middle atmosphere; first, that the amount of ozone is, at all heights, at least roughly in equilibrium with the solar radiation; secondly, that the amount of ozone which is in equilibrium with the solar radiation is but a small fraction of the total amount and is confined mainly in higher altitudes. The ozone present in the lower altitudes is merely that transported to these levels from the region where ozone is actually formed and is thus protected from ozone decomposing solar radiation by the ozone in higher regions.

5. OZONE AS A HEATING AGENT AND THE TEMPERATURE DISTRIBUTION IN THE MIDDLE ATMOSPHERE

We have already mentioned that the high temperature in the middle atmosphere in the region above 35 km., as evidenced by the meteoric phenomena and abnormal propagation of sound waves [Chapters IIIA and IIIB], is due to the absorption of solar radiation by ozone. The ozone layer in fact acts as an enormous heat reservoir raising the temperature of the upper region of the stratosphere to nearly 100°C. The problem of the temperature distribution and of the heat balance in the middle atmosphere due to absorption by ozone has been discussed by Gowan [34*a*, 34*b*] and Penndorf [35] respectively.

Gowan considers the radiation equilibrium of the emitted and absorbed radiation in the various levels of the stratosphere when ozone is mixed with water vapour in different proportions and comes to the conclusion that a rise of temperature will begin from a height of about 30 km. The maximum temperature 400°K. will be attained at a height of about 50 km. Penndorf, on the other hand, considers the absorption and radiation by the ozone layer and shows that in a thin region in the upper part of the ozonosphere the net gain of heat by absorption is about 10 times the loss due to radiation. This thin layer thus acts as an enormous heat reservoir. In what follows we will briefly discuss the results and the methods by which they have been obtained by these two authors.

(*a*) Temperature distribution

Gowan assumes the stratosphere to be non-convective. In such a region any layer of air is dependent for its average temperature on the

solar radiation, on radiation from the solid earth and on radiation from the atmosphere above and below it. The layer is assumed to be homogeneous in the horizontal direction. For computing the radiation equilibrium of the layer, only the vertical radiations are considered. It is further assumed that water vapour and ozone are the principal gases which absorb solar and terrestrial energy. If $k_{\lambda z}$ and $k_{\lambda w}$ be the fractions absorbed at wavelength λ by ozone and water vapour respectively from the radiation passing normally through an elementary layer, the total energy absorbed is given by

$$\int_0^\infty k_{\lambda z}(S_\lambda+s_\lambda+E_\lambda+X_\lambda+Y_\lambda)\,d\lambda$$

$$+\int_0^\infty k_{\lambda w}(S_\lambda+s_\lambda+E_\lambda+X_\lambda+Y_\lambda)\,d\lambda,$$

where S_λ—portion of the solar energy reaching the layer,
s_λ—portion of the solar radiation which, after reflection from the earth's atmosphere, reaches the layer,
E_λ—portion of earth radiation reaching the layer,
X_λ, Y_λ—atmospheric radiation falling on the layer from above and below.

Now the layer emits radiation in directions both upwards and downwards. The effective radiating surface is therefore double the absorbing surface. Hence in the condition of equilibrium we have

$$\int_0^\infty R_{\lambda T_0}\,d\lambda=\frac{1}{2}\int_0^\infty k_{\lambda z}(S_\lambda+s_\lambda+E_\lambda+X_\lambda+Y_\lambda)\,d\lambda$$

$$+\frac{1}{2}\int_0^\infty k_{\lambda w}(S_\lambda+s_\lambda+E_\lambda+X_\lambda+Y_\lambda)\,d\lambda,$$

where $R_{\lambda T_0}$ is the amount of energy radiated by the layer at its temperature T_0 and at wavelength λ and is equal to the energy of a black body at temperature T_0 multiplied by the sum of $k_{\lambda z}$ and $k_{\lambda w}$. On account of the irregular variations of $k_{\lambda z}$ and $k_{\lambda w}$ the above equation can only be solved graphically. Further, the amounts of energy absorbed and radiated by a layer depend primarily upon the amounts of water vapour and ozone present; but the amount of water vapour depends upon the temperature distribution itself. Hence the process of determining the temperature distribution is one of successive approximation. The method is to assume a probable distribution of temperature, determine the proportion of water vapour at various heights on this assumption and, on this basis, calculate what the temperature would be. If the assumed temperature distribution be correct, the final temperatures calculated will agree with the original assumptions. If not, a fresh temperature distribution is assumed, and the process repeated until the calculated temperature distribution agrees with that initially assumed.

In order to carry out calculation by the above method, the stratosphere is divided into nine layers 11 to 15 km. and then every 5 km. up to 55 km. The amount of ozone in different layers is known from experiments of Götz and others [6] and that of water vapour is estimated on the basis of the preliminary estimate of temperature. As a starting point for calculating the amounts of water vapour in different layers it is assumed that at 11 km. height the vapour is saturated at a temperature of 219°K.

To determine the values of $k_{\lambda z}$ and $k_{\lambda w}$ in the spectral region in which selective absorption takes place, we have

$$k_{\lambda z} = \frac{I_0 - I}{I_0} = 1 - 10^{-\alpha_z x}$$

and

$$k_{\lambda w} = 1 - 10^{-\alpha_w x},$$

where I_0 and I are the intensities before and after the radiation has passed through x cm. of the gas at S.T.P. The value of α_z—the absorption coefficient of ozone—is taken from the results of experiments by Fabry and Buisson [36] and by Colange [12]. Knowing α_z and x, $k_{\lambda z}$ is calculated for the spectral ranges 0·23 to 0·34μ, 0·45 to 0·65μ and 1 to 12μ where ozone has selective absorption bands. For water vapour, values of $k_{\lambda w}$ in the spectral range 1 to 50μ are calculated with the help of Hettner's results [37] on absorption coefficients.

So far as the radiation from the layer is concerned, it occurs in the infra-red. This is obtained by plotting the energy of black body radiation corresponding to the temperature of the layer multiplied by $(k_{\lambda z}+k_{\lambda w})$ against λ. The area under this energy curve gives

$$\int_0^\infty R_{\lambda T_0} d\lambda$$

and the integration is carried out from 0 to 50μ beyond which radiation becomes negligible.

The total energy absorbed by a particular layer is obtained after allowing for the amounts which are absorbed in reaching the layer. This is done separately for ozone and for water vapour both for the direct solar radiation *inwards*, and for the terrestrial and reflected solar radiations *outwards*. The terrestrial radiation is assumed to be that of a black body at a temperature of 260°K. For the reflected solar energy it is assumed that 30% of the energy reaching the troposphere is reflected upwards. The amounts of energy corresponding to the above three items are summed up and plotted against λ as one curve for each absorbing agent for each layer. Thus the areas under these curves give the value for ozone

$$\int_0^\infty k_{\lambda z}(S_\lambda + s_\lambda + E_\lambda)\, d\lambda,$$

and for water vapour

$$\int_0^\infty k_{\lambda w}(S_\lambda + s_\lambda + E_\lambda)\, d\lambda.$$

Lastly, the correction for the radiation from the atmosphere itself is considered. Knowing the amount of energy radiated by each layer, we can get the amount absorbed by a particular layer by an analogous method and plot a curve giving energy absorbed against λ. The area under this curve gives the value of

$$\int_0^\infty k_{\lambda(w+z)}\,(X_\lambda+Y_\lambda)\,d\lambda.$$

All the factors required for considering the radiation equilibrium are thus known and the temperature distribution is estimated as mentioned above. The results obtained by Gowan are given in the table below.

It will be seen that in most cases there is a sharp rise of temperature at about 35 km. and a maximum at about 50 km. Moreover, as is to be expected, the maximum temperature attained is greater when there is no water vapour in the stratosphere. It is interesting to compare the calculated temperature distribution of Gowan with that obtained from experiments on

TABLE I

Distribution of temperature in the middle atmosphere calculated (after Gowan) for various distributions of ozone and humidity

Latitude .. Season ..	50° Summer	50° Summer	50° Summer	50° Night	50° Summer (before dawn)
Ozone at S.T.P.	0·28 cm.	0·28 cm.	0·28 cm.	....	0·28 cm.
Relative humidity ..	40%	None	100%	100%	40%
Water vapour distribution	Mixed	....	Mixed	Diffusive equilibrium	Mixed
Height in km.	Abs. Temp.	Abs. Temp.	Abs. Temp.	Abs. Temp.	Abs. Temp.
11–15	225	290	200	205	225
15–20	230	285	195	205	230
20–25	240	285	220	205	239
25–30	245	290	235	210	243
30–35	260	295	245	215	256
35–40	290	335	260	215	276
40–45	370	480	310	215	338
45–50	395	535	315	210	363
50–55	370	600	290	205	347

abnormal propagation of sound waves [see Chapter IIIA]. In the latter case, though the rise of temperature is sharp, yet no maximum is obtained within about 60 km. There is thus a qualitative agreement in so far as the rise of temperature is concerned.

The above results are from Gowan's earlier work [34]. The calculations have been revised by Gowan [34*a*] with newer data on infra-red absorption in ozone (taking pressure into account [34*c*, 34*d*]) and in water-vapour (assuming a square-root law for the pressure affect [34*e*]). It is found that the calculated temperature distribution with height agrees approximately

with the sounding balloon results. It, therefore, seems that the primary assumption that the stratosphere is in radiation equilibrium is justified. The results also explain the existence of high temperatures round 35–50 km. by ozone absorption, though, the ozone maximum is at 22. km.

(b) Heat balance in the middle atmosphere

To discuss the heat balance in the upper stratosphere a knowledge of the influence of radiation (short- and long-wave), of convection, of turbulence and of general circulation upon the temperature of any layer is required. Unfortunately, there is no exact knowledge of the last three processes. The effects of radiation and absorption by ozone alone may therefore be considered after Penndorf [35]. (According to Penndorf water vapour plays an important rôle in the heat preservation of the troposphere only; in the high stratosphere its effect is negligibly small.) In contrast to the previous case radiation equilibrium is not supposed here, but only the net absorption and emission of radiations which may influence the temperature of the stratosphere through the agency of ozone is considered.

Thus we have to calculate and discuss the following:—

(i) Absorption of solar radiation by ozone and the resulting heating of the air.

(ii) Long-wave terrestrial radiation which may also produce heating.

(iii) Characteristic radiation emitted by ozone.

(i) *Absorption of solar radiation.*—Let the solar radiation be incident at angle ζ to the vertical (Fig. 16), $I_{S\lambda}$ be its intensity for wavelength λ at C and BC ($= m$) the equivalent thickness of the ozone layer. Therefore the flow of radiation per unit area in the interval $d\lambda$ passing normally through B in time dt is given by

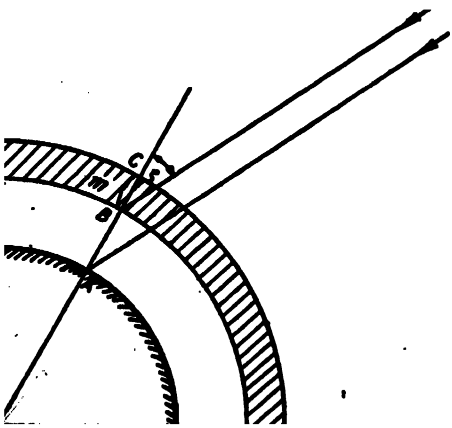


FIG. 16. Absorption of solar radiation incident obliquely on the ozone layer. (After Penndorf.)

$$I_{S\lambda} \cos \zeta \exp [-\alpha_\lambda m/\cos \zeta] \, d\lambda \, dt, \quad \quad .. \quad (14)$$

where α_λ is the coefficient of absorption of ozone for radiation in the wave-length range $\lambda \rightarrow \lambda + d\lambda$. The exponential factor is due to absorption of radiation by the ozone layer.

Since $\cos \zeta = a + b \cos \omega$

where ω—the hour angle of the sun,
$= 2\pi t/\tau$, τ being the unit of time (one day in the present case),
$a = \sin \phi \sin \delta$, $b = \cos \phi \cos \delta$,
ϕ, δ being the latitude of the place and declination of the sun respectively,

the above expression may be written as

$$\frac{\tau}{2\pi} I_{S\lambda} (a+b \cos \omega) \exp\left[-\alpha_\lambda m/(a+b \cos \omega)\right] d\lambda \, d\omega.$$

Therefore, the total flow of solar radiation per unit area passing normally through B in the course of a day is given by

$$S_B = \frac{\tau}{2\pi} \int_\lambda I_{S\lambda}\, d\lambda \int_0^{\omega_0} (a+b \cos \omega) \exp\left[-\alpha_\lambda m/(a+b \cos \omega)\right] d\omega \quad .. \quad (15)$$

where ω_0 is the hour angle for sunrise or sunset. S_B can be evaluated by graphical method. The amount absorbed by the ozone layer in the course of a day is obtained by subtracting (15) from the corresponding flow of the solar radiation into the ozonosphere S_0 in a day and is therefore given by

$$S_E = S_0 - S_B. \quad .. \quad .. \quad .. \quad (16)$$

(ii) *Terrestrial radiation.*—We next consider the absorption of terrestrial radiation by the ozone layer. The earth is considered for our purpose as an infinitely extended black surface. In Fig. 17, f is an element of surface *above* ozonosphere. The solid angle subtended at f by the ring-

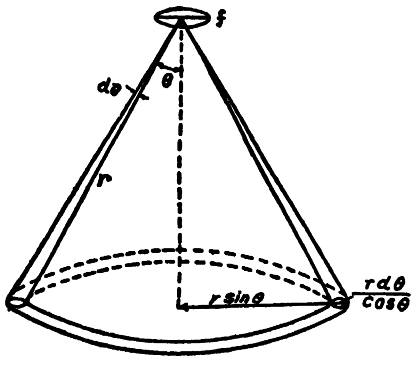


Fig. 17. Absorption of radiation from the surface of the earth by the ozone layer.

shaped area on the surface of the earth is given by

$$2\pi r \sin \theta \cdot \frac{r d\theta}{\cos \theta} \cdot \frac{1}{r^2} = 2\pi \tan \theta \, d\theta.$$

Denoting $I_{E\lambda}d\lambda$ as the amount of radiation from unit area on the earth's surface in the wavelength $\lambda \rightarrow \lambda + d\lambda$, we have the radiation in unit solid angle as $I_{E\lambda}d\lambda/\pi$. Therefore the terrestrial radiation that passes through f from the ring-shaped area after absorption in the ozonosphere is given by

$$\frac{I_{E\lambda} \exp\left[-\alpha_\lambda m/\cos\theta\right] d\lambda}{\pi} \cdot 2\pi \tan\theta \, d\theta \cdot \cos\theta$$
$$= 2I_{E\lambda} \sin\theta \exp\left[-\alpha_\lambda m/\cos\theta\right] d\lambda \, d\theta.$$

We are, however, interested in that fraction of the energy flow of the terrestrial radiation which, moving upwards, intersects the element f perpendicularly. This is obtained by a further multiplication with $\cos\theta$ and is given by

$$2I_{E\lambda} \sin\theta \cos\theta \exp\left[-\alpha_\lambda m/\cos\theta\right] d\lambda \, d\theta.$$

Therefore, the total amount of energy flowing normally through f due to earth's radiation is

$$S_\tau = 2 \int I_{E\lambda} \, d\lambda \int_0^{\pi/2} \sin\theta \cos\theta \exp\left[-\alpha_\lambda m/\cos\theta\right] d\theta. \quad .. \quad (17)$$

Using Gold's H_n-function [38]

$$H_n(\alpha_\lambda \cdot m) = \int^\infty \frac{\exp\left[-\alpha_\lambda m \xi\right]}{\xi^n} d\xi \quad .. \quad .. \quad (18)$$

where $\xi = \dfrac{1}{\cos\theta}$, Eq. (17) may be written as

$$S_\tau = 2 \int_\lambda I_{E\lambda} H_3(\alpha_\lambda \cdot m) \, d\lambda. \quad .. \quad .. \quad .. \quad (19)$$

From the values of the function $H_n\ (\alpha_\lambda \cdot m)$ given in tables of the function of Jahnke-Emde [39], S_T can be determined graphically. From S_T and the flow of terrestrial radiation incident on the ozone layer, the amount absorbed by it can be evaluated.

(iii) *The characteristic radiation of ozone.*—In order to evaluate the total emission of the characteristic radiation from the layer, we can proceed in a manner similar to that for terrestrial radiation. We have only to replace the infinitely extended black surface by an infinitely extended thin layer of ozone as the radiator. In contrast to the black surface, the radiation from the layer is proportional to its thickness and, according to Kirchoff's law, also to the absorption coefficient. Consider a thin layer of ozone of thickness dm at a distance m from f (Fig. 18). The radiation sent in the direction θ from the shaded ring-shaped portion is given by

$$\frac{I_{z\lambda} \exp\left[-\alpha_\lambda m/\cos\theta\right] d\lambda}{\pi} \cdot 2\pi \tan\theta \, d\theta \cdot \frac{dm}{\cos\theta} \cdot \alpha_\lambda \cdot \cos\theta$$
$$= 2I_{z\lambda}\alpha_\lambda \tan\theta \exp\left[-\alpha_\lambda m/\cos\theta\right] d\lambda \, d\theta \, dm, \quad .. \quad (20)$$

where $I_{z\lambda}d\lambda$ is the amount of radiation from unit area of the thin layer of ozone in the wavelength range $\lambda \rightarrow \lambda + d\lambda$. The fraction of this radiation

moving vertically upwards is obtained from (20) by multiplication of $\cos\theta$ and is given by

$$2I_{z\lambda}\alpha_\lambda \sin\theta \exp\left[-\alpha_\lambda m/\cos\theta\right] d\lambda\, d\theta\, dm.$$

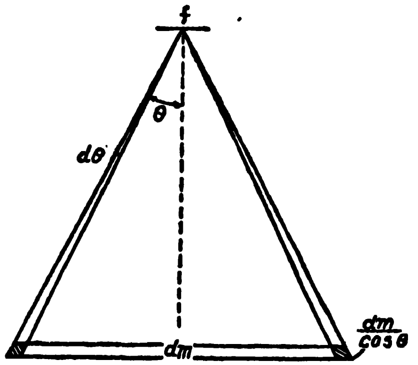


FIG. 18. Emission of radiation from an elementary ozone layer (dm).

Therefore the total amount of energy flowing upwards due to ozone radiation is

$$S_z = 2\int_\lambda I_{z\lambda}\alpha_\lambda d\lambda \int_{m=0}^{m}\int_{\theta=0}^{\pi/2} \sin\theta \exp\left[-\alpha_\lambda m/\cos\theta\right] d\theta dm,$$

which, on introduction of Gold's H_n-function (Eq. (18)), becomes

$$S_z = 2\int_\lambda \int_{m=0}^{m} I_{z\lambda}\alpha_\lambda H_2(\alpha_\lambda . m)\, d\lambda\, dm. \quad .. \quad .. \quad (21)$$

S_z can be evaluated graphically.

The rates of absorption of solar and terrestrial radiation and the rate of emission by the ozone layer are calculated with the help of Eqs. (16), (19) and (21) and from these the rate of heating of the stratosphere is easily obtained from the formula

$$\frac{\partial T}{\partial t} = \frac{\tau}{\rho C_p}\cdot\frac{dQ}{dv\, dt}, \quad .. \quad .. \quad .. \quad .. \quad (22)$$

in which dQ—heat absorbed in volume dv,
ρ—density of air,
C_p—specific heat at constant pressure.

It is to be noted that the temperature change is expressed here in degrees per day per unit volume and not in degrees per day per square centimetre as is usually done. According to Penndorf, the change per unit volume is physically correct because the energy transmission through collision takes place in all directions so that the total energy falling on a surface element distributes itself in the whole volume element.

The assumptions under which the calculations have been made are the following:—

(i) The solar rays are taken as parallel beam. The ultraviolet radiation in the range 2000Å-3000Å is assumed to be that of a black body at temperature 5910°K. Above 3000Å, the value of solar radiation is taken from Abbot's table [1].

(ii) All horizontal gradients of the meteorological elements are neglected.

(iii) The equivalent thickness of the ozone layer and its vertical distribution are taken from the mean curves for Arosa (latitude 45°). The atmosphere from 5 to 47 km. is divided into six layers. Moreover, it is assumed that there is no ozone above 50 km.

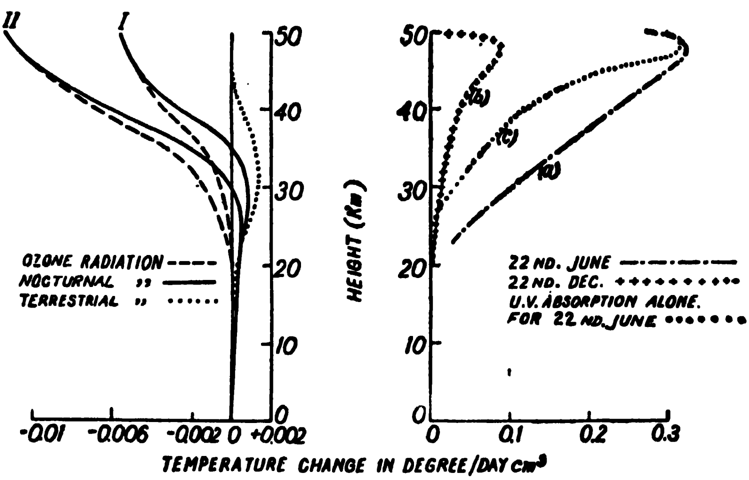
No daily variation of ozone is considered.

(iv) The characteristic radiation and absorption by ozone is taken to be effective in the intervals 0·8 to 11μ and 12 to 17μ.

(v) The temperature of the earth's surface is taken as 280°K. For the temperature of the troposphere at 5 km., 253°K and 233°K are chosen and six different distributions are assumed for the temperature between 5 and 47 km.

Curves in Fig. 20 depict the results of calculation (Eq. (22)). An inspection of the curves shows that there is a reservoir of heat near about 50 km. level which is much above the centre of gravity of ozone. The heating effect is negligible below 20 km. Curve (*c*) shows that the solar ultraviolet radiation responsible for heating (3000Å—2000Å) is almost completely absorbed in a thin layer at the top; the increase in temperature below 25 km. is thus due to absorption in the visible region.

Curves in Fig. 19 show the decrease of temperature per day due to terrestrial and ozone radiations at night. These have been drawn taking



FIGS. 19-20. Illustrating cooling and heating of the middle atmosphere due to ozone absorption. (After Penndorf.)

two different temperature distributions in the stratosphere. Curve I is for a constant temperature up to 50 km. and curve II for the case of rising temperature from tropopause up to 50 km. (280°K at 47 km.). The dotted curve is due to 'negative cooling' on account of terrestrial radiation absorbed in the stratosphere. It is seen that this is effective only within the range 25 to 40 km.

A study of the curves in Figs. 19 and 20 shows that the effect of long- and short-wave radiations is that the loss of heat by nocturnal radiation is very small compared to the gain of heat by absorption of solar energy. In fact, the net gain of heat is at least 10 times greater than the loss. In summer, the gain may be even 60 times greater. It may therefore be concluded that this heat reservoir would not allow the temperature of the stratosphere to remain constant with increasing height.

It may be noted that the height of maximum rise in temperature (50 km.) agrees fairly well with that obtained from the study of the meteoric and the abnormal sound propagation phenomena [Chapters IIIA and IIIB].

Estimates of the cooling of the ozonosphere after sunset have also been made by Gowan [34*b*]. Cooling occurs mainly by radiation up and down and the conclusion is reached that at 50 km. the cooling is about 30°C. but that it is only a fraction of a degree below 25 km. This shows that the temperature inversion in the ozonosphere round 40 km. persists throughout the night.

6. VARIATIONS OF ATMOSPHERIC OZONE—TEMPORAL AND LATITUDINAL

As mentioned in the introduction the ozone content of the atmosphere (i.e. the equivalent thickness) is subject to considerable variations. There are diurnal and seasonal variations, and the average ozone content also depends on the geographical location. The amount over any place also varies considerably with the meteorological conditions. The principal results obtained in respect of these variations are briefly described below.

(*a*) Temporal variations

(i) *Diurnal and nocturnal variations.*—An interesting result of observations on the diurnal and nocturnal variations of ozone is that the average ozone content at night is greater (or, at least never less) than during the day [40, 41, 42, 43]. The average thickness at night is found to be 3·35 mm. No seasonal variation during the period (October, 1927 to April, 1928) of study was observed. During daylight hours there was little, if any, regular hourly variation of ozone. It may, however, be mentioned that such variation has recently been reported from observations made with Dobson apparatus at Delhi (Lat. 28°N.) [43*a*]. It appears that the ozone content diminishes slightly towards noon and then rises again. It has, however, been pointed out by Vassy and Vassy that the observed diminution may not really be a diminution in the thickness of the ozone layer, but a

temperature effect arising out of the author's using the wavelengths λ3110 and λ3300 in his measurements. The absorption coefficients of these two wavelengths have different temperature coefficients and it is shown that a rise in temperature will have the same effect on observational result as a decrease in the thickness of the layer [43b].

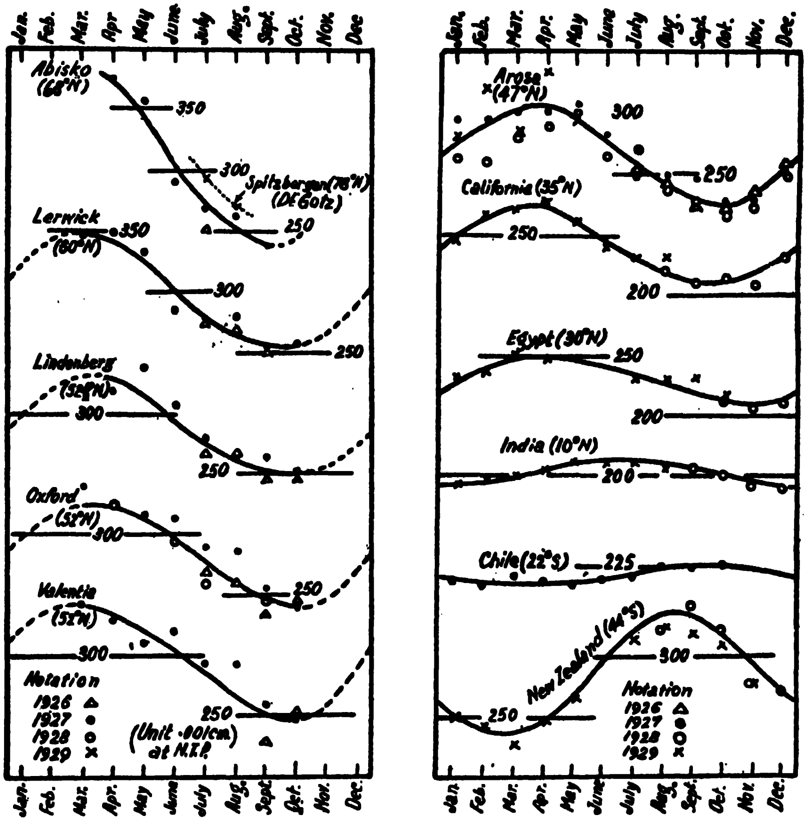


FIG. 21. Seasonal variation of atmospheric ozone content in different latitudes. Note that the amplitude of the variation is least in equatorial regions and highest in high latitudes. The unit is 0·001 cm. at S.T.P. (After Dobson.)

For daylight hours it is difficult to obtain a true mean value on account of minute to minute variation. Compared to this the ozone content during night, as obtained by observation in moon light, is more steady.

(ii) *Seasonal variations.*—The general characteristics of the seasonal variation as observed at different latitudes are shown in Fig. 21 [40]. It will be seen that in general, the ozone content is maximum in spring and minimum in autumn in both the hemispheres. Further, the annual variation is greatest in high latitudes and least at the equator. There are, however, differences in details. Observations on ozone content in subtropical regions have disclosed an interesting feature not seen in the curve referring to India (Kodaikanal, lat. 10°N.). It has been found from observations

at Bombay (19°N.) [44] that besides the usual annual variation typical of northern hemisphere, namely, maximum in April-May and minimum in

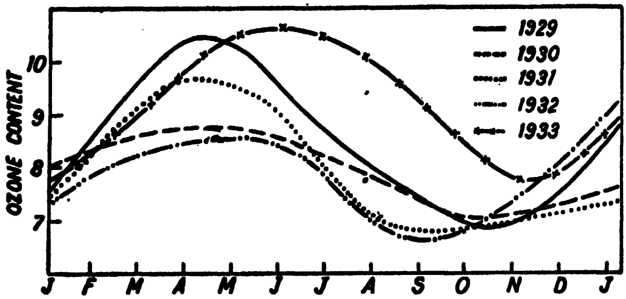


FIG. 22. Yearly variations of ozone content (in arbitrary units) for a number of years at Table Mountain, California. (After Penndorf.)

November-December, there is a variation—abnormally low value—in the monsoon months (July-August) due to some disturbing factor. A similar abnormality has also been observed at Shanghai (31°N.) [45]. A study of the seasonal variations made at Calcutta (22° 33′ N.) [46] by the radio method [Chapter VI, Sec. 12(*d*)] revealed distinctly this abnormality.

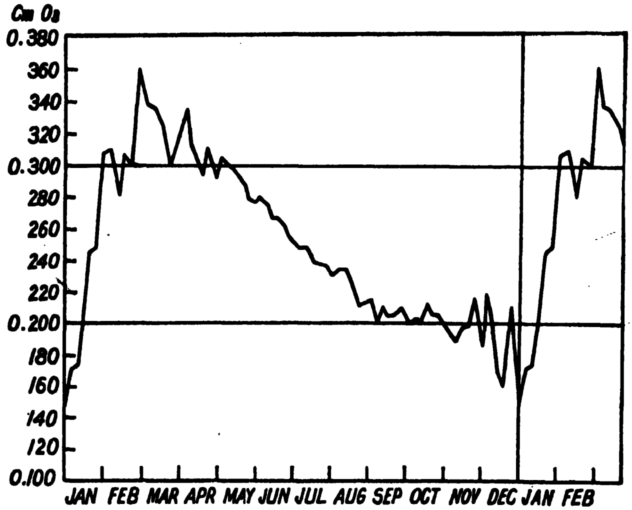


FIG. 23. Mean yearly variation of the ozone content at Tromsö. The curve depicts the means for each *pentad* derived from the normal daily values calculated from observational data of the period 1926-42.

Figs. 22 and 23 depict yearly variations for a number of years as observed at Table Mountain (lat. 35°N.) and at Tromsö (lat. 70° N.) [35, 35(*a*)]. The latter station is interesting because of long winter nights. The curve for this station (Fig. 23) depicts the means for each *pentad* derived from the normal daily values calculated from observational data of the period 1926–1942. It will be noticed that the annual variation does not

follow a simple sine-curve and also that there is much larger interdiurnal variation in the winter than in the summer values. According to Penndorf [35(a)] these characteristics are explained if it is remembered that in the polar latitudes no ozone is formed during the months when solar radiation is missing (polar night) or when the sun remains above the horizon, at very low angle, for a few hours only. There is thus no photo-chemical equilibrium of ozone during this period. The interdiurnal variations are then solely due to advection. If there is a flow of air from arctic regions then the ozone value is low; if, however, the flow is from southern latitudes then the air is richer in ozone. A study of the ozone content in the polar regions may thus lead to information concerning general circulation of the lower stratosphere over these regions in winter.

(*b*) Latitudinal variation

Fig. 24 shows the geographical distribution of ozone-values for different months of the year [40]. It will be seen that the amount of ozone near the equator is small and fairly constant, the average value being about 2·0 mm.

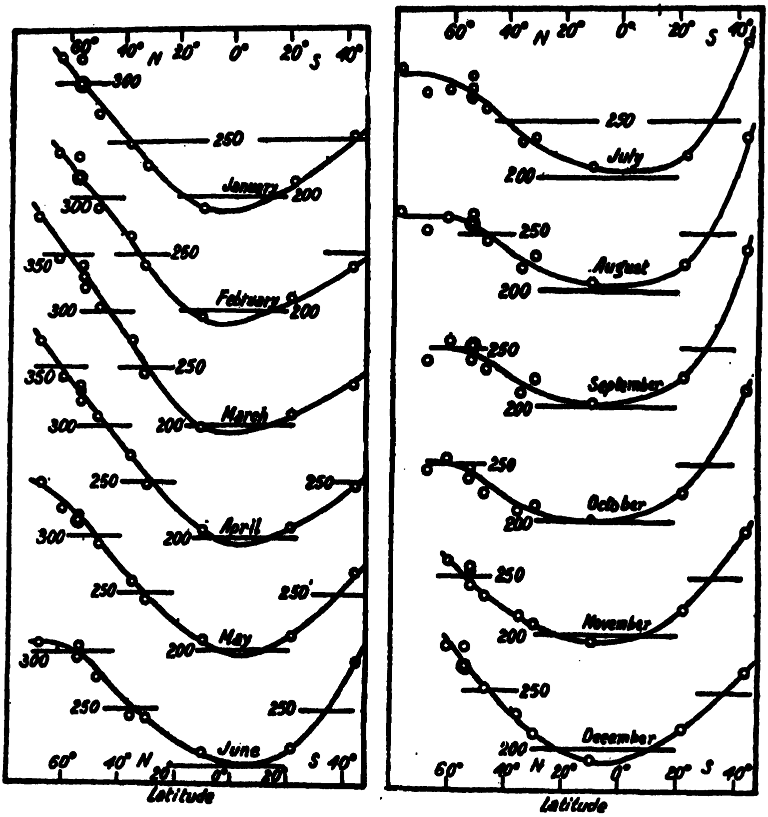


FIG. 24. Geographical distribution of ozone-value for different months of the year. The unit is 0·001 cm. at S.T.P. (After Dobson.)

In high latitudes there is large annual variation. It is very large in spring but in autumn, the increase is not more than 25 per cent at the equator.

The results of observations described above clearly show that the ozone content is markedly influenced by the variation of the solar ultra-violet radiation. Since solar radiation both produces and destroys ozone, it is evident that the increase or the decrease of the ozone-value with the inclination of the incident radiation will depend upon the relative intensities of these two processes. Starting with this idea Chapman [47] has studied theoretically the variation of ozone content of the atmosphere. His analysis shows that the variations of the ozone content ought to have the observed form if due account is taken of the processes of production and destruction of ozone occurring simultaneously. The agreement, however, is rough and only qualitative due to various simplifying assumptions made.

More definite suggestion regarding the cause of the low value of ozone in the equatorial regions has been made by Malurkar [47*a*]. Measurements show that the amount of water vapour above the troposphere is greater in the equatorial regions than over the corresponding levels in temperate latitudes. Further, the amount of solar radiation falling on unit area is greater in the equatorial than in the higher latitudes. Hence a reduction in ozone content in equatorial regions is caused by increased photolysis of O_3 in these regions by the solar radiation in presence of water vapour [see Sec. 4(*a*)]. It is interesting to note in this connection that the presence of the OH radicle in the high atmosphere has been proved by the occurrence of strong OH bands in the *infra-red* in the night sky spectrum (see Chap. X).

7. OZONE AND WEATHER

It has now been established that definite correlation exists between weather condition in the troposphere and ozone content in the stratosphere. The following remarks are from observations carried out by Dobson and his associates [48].

The passage of a warm front is accompanied by a fall in the ozone content, while that of a cold front by a rise in the ozone content. These effects are, however, observed only when the fronts extend well into the stratosphere.

A warm front in passing over the ground surface replaces cold air by warm air. As the front slopes forward at an angle of about 1/150, the fall in the ozone content takes place long before the arrival of the front at the ground surface. This is analogous to the appearance of upper clouds which herald the approach of a warm front at the ground level. In Fig. 25 the fall in the ozone content in the case of an advancing warm front as observed in England on November 17–19, 1940, is depicted. The figure is a cross section of the atmosphere at right angles to the direction of the front. The region where the most rapid fall of ozone occurred is shown by the line at the top. It will be noted that the fall occurred when the height of the front overhead was about 8 km. It is, however, believed that

the fall was associated directly with the extension of the front in the stratosphere (with a backward slope) rather than with its position in the tro-

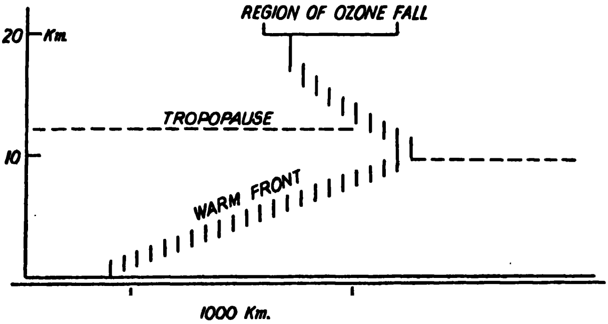


FIG. 25. Illustrating the fall in the ozone content in the case of an advancing warm front as observed in England on November 17–19, 1940. The figure is a section of the atmosphere at right angles to the direction of the warm front. (After Dobson, Brewer and Cwilong.)

posphere. This type of fall in ozone content associated with the passage of a warm front is extremely common. As a matter of fact, the first sign that can be obtained from ground observations of the approach of a warm front, is a drop in ozone value. There are some cases, however, in which a change in the ozone value is not observed. For these the fronts do not presumably extend into the stratosphere.

As expected, the changes associated with a cold front are reverse to those for a warm front. There is generally a rise in the ozone content associated with a cold front. A cold front slopes backward, but as the slope is much steeper than that of a warm front, the associated rise in ozone content occurs soon after the front passes over the ground surface. As in the case of warm fronts, all cold fronts do not show a change in the overhead ozone content. This is explained as due to the front not extending up into the stratosphere. Fig. 26 is a section of the atmosphere at right

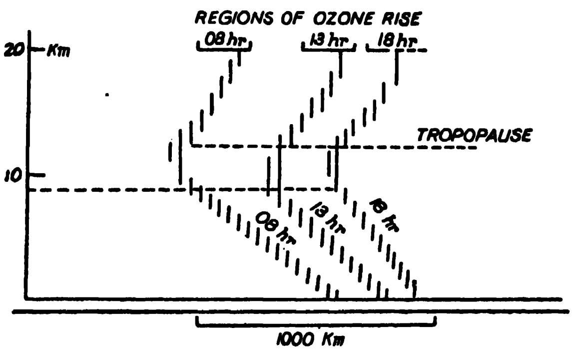


FIG. 26. Illustrating the fall in the ozone content in the case of an advancing cold front as observed in England on June 24-25, 1940. The figure is a section of the atmosphere at right angles to the cold front. (After Dobson, Brewer and Cwilong.)

angles to a cold front which passed over England on June 24-25, 1940. The height of the tropopause as shown is inferred from observations made some distance away from the front. The position of the front in the upper troposphere was fixed by meteorological data as were available. The position in the stratosphere is determined by consideration of the increase of the ozone content.

Some interesting observations on the change of ozone content have been made when an occlusion passes overhead. For the case of true occlusion there is always a fall in the ozone content. (A true occlusion is one in which the warm sector of the depression continues aloft extending well ahead—several hundreds of kilometres—of the tip of the warm sector at the ground level.) There are, however, also cases of occlusion in which no fall in ozone content occurs or in which a marked rise is observed. In the first of these cases the occlusion presumably does not extend into the stratosphere. For the second case, the warm sector with which the occlusion is associated may be a great distance away from the centre of the depression. In all such cases the air over the occlusion seems to be cold rather than warm.

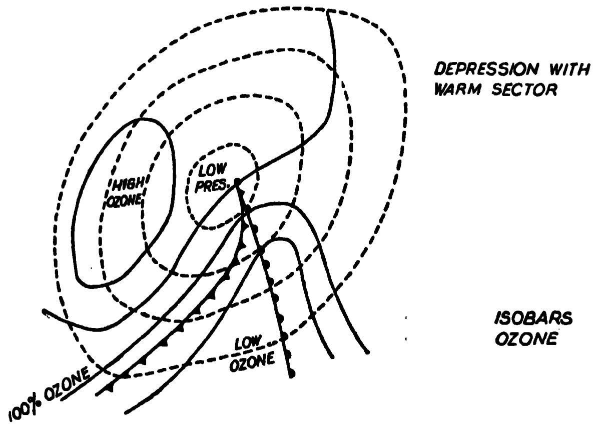


FIG. 27. Illustrating the distribution of ozone around a typical young depression with a very marked warm sector. Note that the ozone content is low over the warm sector and is high above the advancing cold air mass behind. (After Dobson, Brewer and Cwilong.)

Fig. 27 shows after Dobson *et al.* [48] the distribution of ozone around a typical young depression with a well marked warm sector. It will be noticed that the ozone content is low over the warm sector and is high above the advancing cold air mass behind.

From what has been said above, high ozone value is to be expected with cyclonic systems (low pressure) and low value with anti-cyclonic system (high pressure). This is depicted in Figs. 28(*a*) and (*b*). The continuous lines are isobars. The ozone-values as observed for each normal cyclone and anticyclone that occurred are plotted on these maps in their appropriate

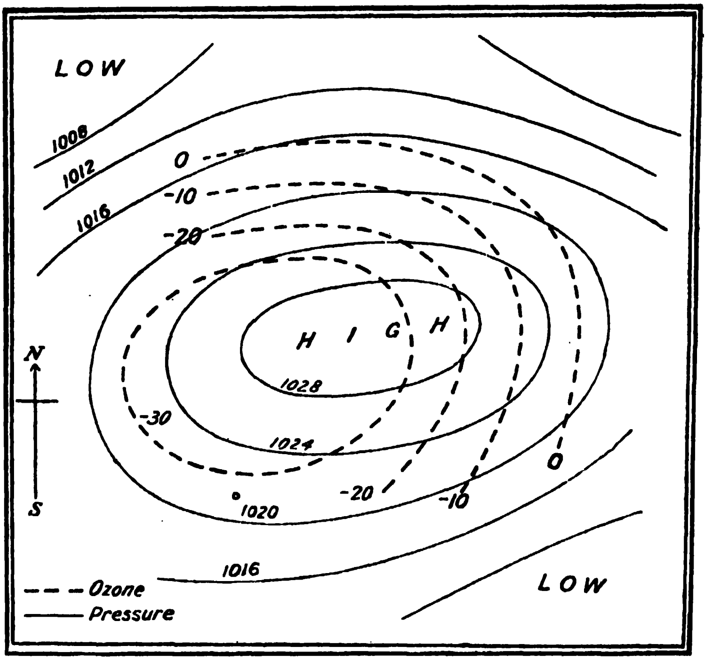


FIG. 28(a)

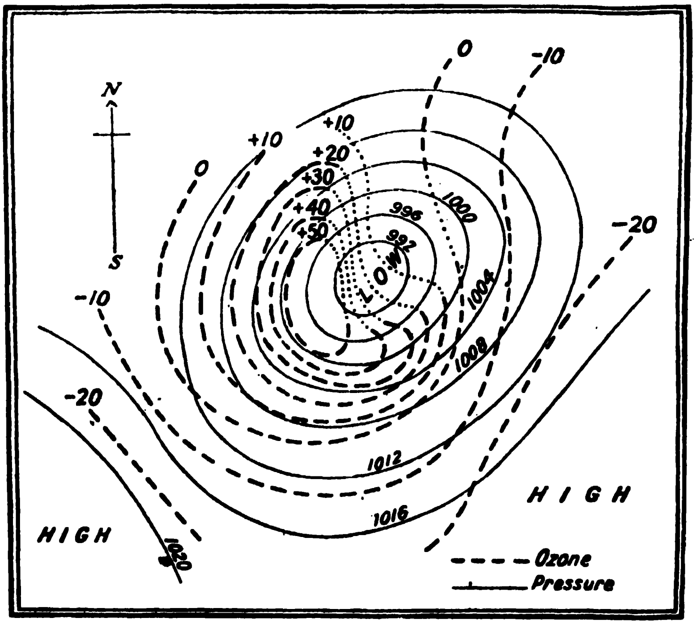


FIG. 28(b)

FIGS. 28(a, b). Illustrating how the amount of ozone is related to pressure distribution in cyclone and anticyclone. Continuous lines are isobars; the dashed curves are lines of equal ozone content. The numbers indicate departures from the monthly ozone-values. They are positive or negative according as the observed value is greater or less than the monthly mean value. (After Dobson, Harrison and Lawrence.)

positions according to the pressure distribution. The broken line curves are drawn through these points. Since the ozone content is subject to seasonal variation, the differences from the monthly ozone-values are used. The values plotted are thus positive or negative according as the ozone content during the cyclone is greater or less than the mean value of the month in which the cyclone has occurred.

In some later observations at Oxford (July 1941) it was found that when thunderstorms passed overhead there was a rise of 2 mm. in the ozone content (from 2·7 to 4·7 mm.) which lasted from one to three or four hours.

Amongst other observations on ozone content and the weather, mention may be made of those of Tönsberg and Chalonge [49] at Tromsö (lat. 70°N.) who found that there is a close correlation between the two. Fig. 29 depicting the day-to-day ozone-values for one year shows that the periods of fine and settled weather (marked F.W. in the figure) are associated with the relatively lowest and most constant ozone-values. The scatter of the daily ozone-values is greatest in spring and is associated with relatively bad and variable weather conditions. The large jumps in ozone content occurring in this season may be attributed to violent changes in the atmosphere associated with extremely deep cyclones in the Atlantic Ocean. It is further found that the tropical air has a low, the polar air a medium, and the arctic air the highest ozone content.

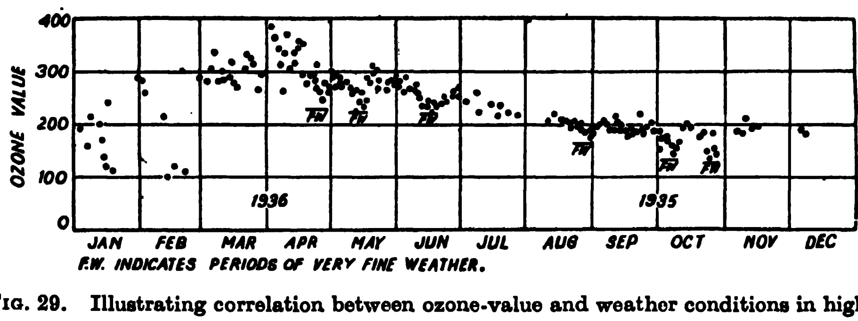


FIG. 29. Illustrating correlation between ozone-value and weather conditions in high latitude (Tromsö, 70°N.). The unit is 0·001 cm. at S.T.P. (After Tönsberg and Chalonge.)

(b) Pressure and temperature in the troposphere

It has been shown by Dobson [40] that in temperate latitudes, the difference in the amount of ozone from its monthly mean is closely correlated negatively with the pressure at 9 to 16 km. level and with the mean temperature of the troposphere (Fig. 30).

Meetham [50], after a careful analysis of available data of ozone-values and the meteorological data, has shown that ozone has a surprisingly high correlation with the potential temperature and the density in the stratosphere.

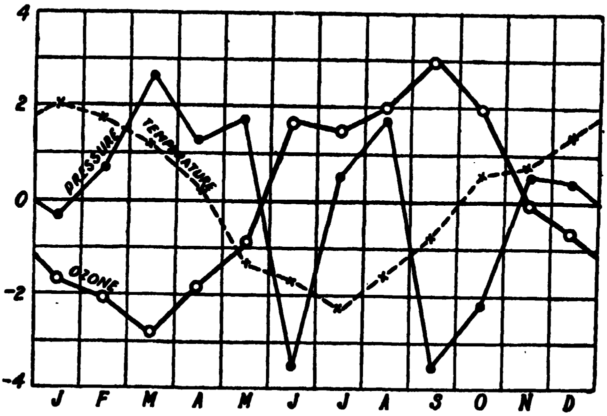


FIG. 30. Illustrating the negative correlation between the variation in ozone content and the pressure at 9–16 km. level and also with the mean temperature of the troposphere (1 scale div. = 0·02 of O_3, 0·1 inch, 5°F.). (After Dobson.)

8. OZONE AND SOLAR CYCLE

Cabannes and Dufay [51] announced in 1927 that they had found some correlation between sunspot activity and ozone content. That there might be some connection between the two was also suggested by Fowle in 1929 [52]. Fowle's later work in 1934 does not, however, support this view [53] More recent observations also seem to indicate that the average variation of the ozone content is not associated with the eleven-year solar cycle [54]. It appears, however, that ozone content variation has 27 and 15·5 days periods with small amplitudes [55].

Dobson, Harrison and Lawrence [48] have shown that there is a definite relation between variation of the ozone content and terrestrial magnetic disturbances. An increase in the ozone-value in the middle atmosphere is always associated with magnetic storms.

9. EFFECTS OF OZONE 'SHADOW'

It was pointed out by Mitra [56] and also independently by Götz [57] that the absorption by the ozone layer increases the radius of the earth's shadow for the ultraviolet light of the sun and that the amount of the increase is equal to the altitude of the upper boundary of the ozone layer.

Penndorf [58] has discussed in detail a number of effects due to this screening and has, for the purpose, introduced the useful terms *ozone shadow* and *boundary of ozone shadow*. Ozone shadow for radiation of a given wavelength is defined as the region in space which is so screened by the earth (together with the ozonosphere) that the intensity of solar radiation for the chosen wavelength λ is *less* than one per cent of the incident intensity. The boundary of the ozone shadow is defined as the spatial curve, at all points of which the intensity of the wavelength λ is *exactly equal* to 1 per cent of the incident intensity. It is found that for a ray passing

tangentially to the ozonosphere, only a thin layer at the top suffices, on account of the strong absorption, to reduce the intensity in the wavelength range $\lambda 2150$ to $\lambda 2900$ to 1 per cent. Assuming that the upper boundary of the ozone layer is at 50 km. height and that the ozone is in photochemical equilibrium in the region 40 km. to 50 km. and is distributed accordingly, Penndorf finds the thickness of this top layer to be less than 1 km.

The following effects of the ozone shadow may be mentioned in particular: (1) Delay in the hour at which ionization in the upper atmosphere (*E*-layer) begins to increase at sunrise. This has been studied by Ghosh [62] and also by Penndorf [59] and will be discussed in Chapter VI, Sec. 12(*e*). (2) The premature decay in the intensity of the sodium *D*-lines from the upper atmosphere at sunset. This phenomenon may be utilized for estimating the upper boundary of the ozone layer and is referred to in Chapter X, Sec. 9. (3) Variation in the intensity of the ultraviolet light reflected from the moon during lunar eclipse. This effect, properly utilized, enables one to determine the vertical distribution of ozone as described below. Detailed discussion on all the three effects are given by Penndorf [58, 59]. (4) To complete the list, one may also add an effect observed by Störmer [60] in respect of the increase in the luminescence of noctilucent clouds at sunrise. This increase begins after the clouds are struck by solar rays passing 30–45 km. above the earth, i.e. above the main bulk of ozone.

The method of utilizing lunar eclipse for determination of ozone distribution was suggested by Götz [57] and first experimentally carried out by Barbier, Chalonge and Vigroux [61].

During a lunar eclipse the 'optical' shadow of the earth proper on the moon is surrounded by the ozone shadow. It is clear that as one passes out radially of the optical shadow, one finds that the light incident on the moon's surface at different radial distances has travelled through different masses of ozone in the ozonosphere at different heights. Measurement of absorption of light, received from points situated at different radial distances from the edge of the optical shadow of the earth, thus enables one (knowing the absorption coefficient) to determine the mass of ozone in the different ray paths at different heights above the surface of the earth. From these measurements on ozone mass as function of ray paths at different heights, it is possible to obtain ozone content as a function of height. It is to be noted that the workers mentioned above utilized, instead of ultraviolet absorption, absorption in the Chappuis bands in the visible part of the spectrum. The results obtained agree well with those deduced from Umkehr effects.

The method suffers from the obvious disadvantage that the measurement can be made only during an eclipse. But it has the advantage that whenever such measurement is possible it provides more reliable ozone data for the level above the maximum (between 40 and 45 km.) at which other methods of measurement are difficult. (See also ref. [63].)

CHAPTER V

OXYGEN AND NITROGEN IN THE UPPER ATMOSPHERE

1. INTRODUCTION

The auroral and the night sky air-glow spectra provide indisputable spectroscopic evidence of the presence of oxygen and nitrogen in the atomic state in the upper atmosphere. (See Chap. VIII, Sec. 6 and Chap. X, Sec. 4(c). For spectroscopic notations, see Appendix, Sec. 2.)

Atomic oxygen is produced by the dissociative action of the solar ultraviolet radiation in the Runge-Schumann continuum (λλ1751–1200) on O_2-molecules

$$O_2 + h\nu \rightarrow O(^1P) + O(^1D) \qquad \ldots \qquad \ldots \qquad \ldots \qquad (1)$$

The absorption is extremely strong. In the region λ1450, where it is strongest, a thickness of 0·0014 cm. at S.T.P. reduces its intensity to one-half. Since the amount of O_2 above 80 km. is much greater than this amount, it is clear that the whole of the radiation will be absorbed above this level and that the oxygen in the atmosphere above will be mainly in the atomic state.

In contrast to this, the mode of production of nitrogen atoms in the upper atmosphere is rather involved, there being no known simple photo-dissociation process analogous to that in reaction (1). One is also not sure if the nitrogen above a certain level, like oxygen, is almost wholly dissociated. In fact, our knowledge of the distribution of atomic nitrogen in the upper atmosphere is still meagre compared to that of atomic oxygen, though, there is no doubt about its presence.

In what follows we shall first give an account of the theoretical work and the results that have been obtained thereof on the height distribution of atomic oxygen in the upper atmosphere. A short discussion on the processes which may lead to the dissociation of N_2 and on the probable distribution of nitrogen atoms in the upper atmosphere as suggested by the various workers will then follow.

2. DISTRIBUTION OF ATOMIC OXYGEN IN THE UPPER ATMOSPHERE

(a) Recombination processes

In the state of equilibrium distribution of the O-atoms there is balance between two opposing processes, viz. dissociation of the O_2-molecules and the recombination of the O-atoms to re-form the O_2-molecules. Hence, to calculate the equilibrium distribution we have to consider, besides the dissociation of the molecules according to the reaction (1), the possible recombination processes which may be effective in maintaining the

equilibrium. The two main recombination processes by which the O-atoms can recombine are, (1) recombination by two-body collision (radiative recombination) and (2) recombination by three-body collision [see Chap. VI, Sec. 12(*a*)].

Radiative Recombination—(two-body collision).—Let both the oxygen atoms be in the normal ^3P state. Then we have

$$O(^3P)+O(^3P) \longrightarrow O_2^*+h\nu \quad .. \quad .. \quad .. \quad (2)$$

The energy released on recombination is used up partly in exciting the O_2 molecule and is partly carried away by the quantum of radiation $h\nu$. The O_2 molecule might be excited to one of the states appropriate for emitting the telluric bands or to one of the states for emitting the Herzberg bands [see Appendix, Sec. (2)]. The probability of recombination in the former state is known to be several orders smaller than that in the latter state. Hence we consider only the process in which the molecule formed is excited to emit the Herzberg bands. If the concentration of the ^3P oxygen atoms be denoted by $[O(^3P)]$, then the number of atoms disappearing by recombinations per unit volume per unit time is given by $\alpha[O(^3P)]^2$, where α is the recombination coefficient. Unfortunately, the value of α is not known with certainty. According to Herzberg [1] its value is of the order of 10^{-16} cm.^{3}/sec. According to Nicolet's [2] estimate, however, it is only 10^{-20} cm.^{3}/sec.

Since, in the processes of photo-dissociation O-atoms in ^1D state are also produced, we should also take account of the recombination process which is reverse of the photo-dissociation process, namely,

$$O(^3P)+O(^1D) \longrightarrow O_2^*+h\nu. \quad .. \quad .. \quad .. \quad (3)$$

The probability of this reaction is many orders higher than (2) and the coefficient of recombination is estimated to be 2×10^{-14} cm.^{3}/sec. [2]. But the effectiveness of the process naturally depends on the number of ^1D atoms in comparison with the number of ^3P atoms.

Three-body collision.—In this case we have

$$O+O+M \longrightarrow O_2^*+M$$

where M is the third body which takes away the excess of energy and momentum. If the concentrations of O and M are denoted by [O] and $[M]$ respectively, then the number of O-atoms disappearing in unit volume per sec. by the process is $\beta[O]^2[M]$. The value of β is estimated to be of the order of 10^{-32} cm.^{6}/sec. [2, 3, 4].

With the above values of the recombination coefficients we can determine the pressure (or, the number density of particles) at which recombination by three-body collision will begin to preponderate over recombination by two-body collision.

Let $[N]$ be the concentration of the colliding particles of all kinds and let $K[N]$, the fraction of $[N]$, be the concentration of atomic oxygen. Then $[M]$, the concentration of the third body, is $(1-K)[N]$.

We then have for the rate of recombination by three-body collision

$$K^2[N]^2(1-K)[N]\,\beta = K^2(1-K)[N]^3\beta.$$

For the rate of recombination by two-body collision we have $K^2[N]^2\,\alpha$, where α is the recombination coefficient. Hence,

$$\frac{\text{recombination rate for 3-body process}}{\text{recombination rate for 2-body process}} = \frac{K^2(1-K)[N]^3\beta}{K^2[N]^2\alpha} = (1-K)[N]\,\beta/\alpha.$$

As stated above, β is estimated to be 10^{-32} cm.^{6}/sec. If the value of α is taken as 10^{-16} cm.^{3}/sec. after Herzberg [1], the above ratio is found to be $(1-K)[N]\,10^{-16}$. Hence, if the concentration of oxygen atoms be small compared to that of the third body, that is if $K \ll 1$ (condition most favourable for preponderance of 3-body process over 2-body process), we have, for the ratio of the rates, $[N]\,10^{-16}$. This will be greater than unity, that is, three-body rate will be greater than two-body rate if $[N] > 10^{16}$/cm.^{3} This atmospheric density prevails at height of about 55 km. But, as already mentioned,§ the dissociating radiation $\lambda\lambda$ 1751–1200 is totally absorbed much higher up. Hence, if the value of α is taken to be of the order of 10^{-16} cm.^{3}/sec., radiative recombination by two-body collision will be the process determining the distribution of O-atoms in the upper atmosphere.

If, however, the value of α is taken as $7{\cdot}6\times 10^{-20}$ cm.^{3}/sec. after Nicolet [2], then we find that the atmospheric density at which three-body collision begins to predominate is $7{\cdot}6\times 10^{12}$/cm.^{3} This density prevails at the height of 110 km. and is within the region of absorption of the dissociating radiation. If, therefore, one adopts this latter value of α, one has to take into account, at least for the base of the atomic oxygen region, the disappearance of the O-atoms by the three-body collision process.

It is to be noted that there is a third process, besides the two discussed above, by which the O-atoms may disappear. The atmospheric region where atomic oxygen predominates is also rich in ionization. Since atomic oxygen has considerable electron affinity we would expect copious formation of negative ions of atomic oxygen and the following reaction is possible:

$$O^- + O \longrightarrow O_2 + e \quad \ldots \quad \ldots \quad \ldots \quad (4)$$

Here the attached electron acts as the third body carrying away the extra energy and momentum. The probability of this reaction has been examined by Massey [5] and is found to be of the same order as that of radiative recombination of O-atoms. Loss of O-atoms due to this reaction, however, depends upon the concentration of O^- ions.

Distribution of atomic oxygen in the upper atmosphere has been computed by different authors by taking either the two-body or the three-body process as the principal recombination process. The results generally agree in that the region of atomic oxygen is found to lie almost entirely above 90 km. The actual distribution curve and the position of the maximum concentration differ, however, by significant amounts from the work of one author to that of the others.

The problem of the photo-dissociation of the O_2-molecules and the distribution of the resulting O-atoms in the upper atmosphere is very similar to that of the photo-ionization of the atmospheric gases and the distribution of the resulting electrons and ions in the upper atmosphere. The latter problem has been extensively studied in connection with that of the formation and stratification of the ionosphere and is discussed in Chapter VI, Sec. 10(*d*). The problem of O_2 has also been studied by analogous methods. In what follows we will first describe in some detail an elegant method in which an extension of Saha's well-known theory of thermal ionization [6] is adapted to the present problem of dissociation. Results of studies by other methods will be given briefly afterwards.

(*b*) Extension of the thermal dissociation method

As mentioned above, the extension of thermal dissociation method is based on the well-known theory of thermal ionization of Saha [6]. Thermal ionization is regarded as a case of reversible chemical reaction so that the percentage of molecules ionized in a gas in an enclosure at a given pressure and temperature can be calculated in just the same way as in homogeneous chemical reaction. In this case the enclosed radiation at temperature T is obviously in thermodynamic equilibrium with the ionization products. But we can also have the case where the radiation is at a temperature higher than that fixed by the thermodynamical equilibrium condition. This latter condition prevails in stellar atmosphere, where the comparatively cool envelop of gas is traversed by radiation at much higher temperature from the hot body of the star inside. The theory of thermal ionization has been extended by Woltjer [7] to such cases. This extension of the theory of thermal ionization had been applied by Pannekoek [8] to study ionization distribution of the terrestrial atmosphere which is traversed by hot radiation from the sun. Since the case of dissociation of molecules is analogous to that of ionization, the method of Pannekoek may also be applied to the study of dissociation of O_2-molecules. This was first done by Majumdar [9] who obtained the distribution of O-atoms in the upper atmosphere. In what follows we will describe this method in some detail after the more recent work of Rakshit [10] on the subject. (See also ref. [10*a*].)

(i) *Case of thermodynamic equilibrium.*—Consider first the case when the gas and radiation are at the same temperature in thermodynamic equilibrium. The equilibrium condition is given by the so-called *reaction isochore.*

$$\frac{n_A n_B}{n_{AB}} = \frac{Gs}{hr_0^2}\left(\frac{MkT}{8\pi}\right)^{\frac{1}{2}} e^{-\frac{D}{kT}}\left(1-e^{-\frac{h\omega}{kT}}\right), \quad .. \quad .. \quad (5)$$

where n_A, n_B, n_{AB} are respectively the numbers per unit volume of free atoms A and B and the molecule AB,

$G = G_A G_B / G_{AB}$,

G_A, G_B—the statistical weights of the normal quantum states of the free atoms,

G_{AB}—the statistical weight of the ground state of the electronic configuration of the molecule, the nuclei being regarded as fixed,

s—the symmetry number of the molecule and is equal to 2 when the two atoms produced by dissociation are identical; when the atoms are not so, $s = 1$,

r_0—the mean separation between the atoms,

M—the reduced mass of the molecule and given by

$$M = \frac{m_A m_B}{m_A + m_B},$$

m_A, m_B being the masses of the corresponding atoms,

D—the energy of dissociation of the molecule AB into its constituent atoms A and B,

ω—the fundamental vibration frequency of the molecule.

From the above relation, knowing the values of the various atomic constants, one can calculate the proportion of molecules dissociated for any pressure and temperature. If one of the two dissociated atoms, say A, be in an excited state A', then for the relative numbers of the excited and unexcited atoms we have, from Boltzmann's law,

$$\frac{n_{A'}}{n_A} = \frac{G_{A'}}{G_A} e^{-\frac{E}{kT}},$$

where E is the energy of excitation of the atom A. This is justified in cases where the lifetime of the excited atoms is large compared with the time between successive atomic collisions. Substituting in (5)

$$\left[\frac{n_{A'} n_B}{n_{AB}}\right]_0 = \frac{G's}{hr_0^2}\left(\frac{MkT}{8\pi}\right)^{\frac{1}{2}} e^{-\frac{D+E}{kT}}\left(1-e^{-\frac{h\omega}{kT}}\right) \qquad .. \quad (6)$$

where the suffix (0) refers to the case of thermodynamic equilibrium;

$$G' = G_{A'} G_B / G_{AB}.$$

It will be noted that Eqs. (5) and (6) giving the degree of dissociation of the molecules are derived from the condition that in the equilibrium state the rate at which the molecules dissociate into atoms is equal to the rate at which they are formed by recombination. They do not make any reference to the details of the mechanism by which the equilibrium is maintained.

These relations may also be obtained by taking into account the details of the absorption processes which lead to dissociation and the collision processes which lead to recombination.

Dissociation.—Let the threshold frequency for dissociation be ν_0. Then the number of absorption processes per second per unit volume due to isotropic radiation within the frequency range ν and $\nu + d\nu$ ($\nu > \nu_0$) which leads to dissociation is

$$n_{AB}\psi_\nu c\rho_\nu d\nu / h\nu, \qquad .. \quad .. \quad .. \quad .. \quad (7)$$

where ψ_ν is the molecular absorption coefficient and ρ_ν is the density of radiation of frequency ν.

The dissociated atoms A and B will separate with a relative velocity V given by

$$\tfrac{1}{2}MV^2 = h\nu - h\nu_0, \qquad \ldots \quad \ldots \quad \ldots \quad (8)$$

where $h\nu_0 = D$, the energy of dissociation. If one of the dissociated atoms be in an excited state and E be the energy of excitation, then $h\nu_0 = D+E$.

Recombination.—The atoms thus formed will, in turn, recombine to form back the original molecule AB. We assume that the gas absorbing radiation is raised in temperature and that in spite of the fact that the dissociated atoms are ejected with velocities according to Eq. (8), the velocities of the particles settle down to Maxwellian distribution corresponding to the equilibrium temperature of the gas.

Now the total number of collisions of the atoms with relative velocity between V and $V+dV$ is

$$\alpha \,.\, 4\pi n_A . n_B \left(\frac{M}{2\pi kT}\right)^{\frac{3}{2}} e^{-\frac{MV^2}{2kT}} V^3\, dV,$$

where α is a coefficient having the dimension of an area, which is a function solely of the atoms and the relative velocity with which they collide. But every one of these collisions does not result in recombination. And, the total number of spontaneous recombinations, due to collisions, per unit volume per second may be written as

$$4\pi n_A . n_B \beta_\nu \left(\frac{M}{2\pi kT}\right)^{\frac{3}{2}} e^{-\frac{MV^2}{2kT}} V^3\, dV, \qquad \ldots \quad \ldots \quad (9)$$

where β_ν is a quantity proportional to ψ_ν. (See below.)

Besides spontaneous recombination, there may be recombinations stimulated by the radiation field, given, as usual, by

$$\frac{\text{stimulated recombinations}}{\text{spontaneous recombinations}} = \frac{c^3\rho_\nu}{8\pi h\nu^3}.$$

Therefore, the grand total number of recombinations is

$$4\pi n_A . n_B \beta_\nu \left(\frac{M}{2\pi kT}\right)^{\frac{3}{2}} \left(1+\frac{c^3\rho_\nu}{8\pi h\nu^3}\right) e^{-\frac{MV^2}{2kT}} V^3\, dV. \qquad \ldots \quad (10)$$

The expression for β_ν is easily obtained from the ratio $n_A . n_B / n_{AB}$ as given in (6), since (7) and (10) should be equal for any frequency $\nu > \nu_0$. Thus,

$$\frac{G's}{hr_0^2}\left(\frac{MkT}{8\pi}\right)^{\frac{1}{2}} e^{-\frac{h\nu_0}{kT}} \left(1-e^{-\frac{h\omega}{kT}}\right)$$

$$= \frac{\psi_\nu}{\beta_\nu} \cdot \frac{c\rho_\nu}{4\pi h\nu}\left(\frac{M}{2\pi kT}\right)^{-\frac{3}{2}} \left(1+\frac{c^3\rho_\nu}{8\pi h\nu^3}\right)^{-1} e^{-\frac{MV^2}{2kT}} \frac{M}{hV^2} \quad \ldots \quad (11)$$

because $MV\,dV = hd\nu$. For isotropic radiation,

$$\rho_\nu = \frac{8\pi h\nu^3}{c^3} \cdot \frac{1}{e^{h\nu/kT}-1}.$$

Substituting in (11) we get on simplification

$$\beta_\nu = \frac{8\pi^2 r_0^2}{c^2 G's} \cdot \frac{\psi_\nu kT}{\frac{1}{2}MV^2} \cdot \frac{\nu^2}{\left(1-e^{-h\omega/kT}\right)} \cdot \qquad \ldots \qquad \ldots \qquad (12)$$

(ii) *Case of radiation at higher temperature.*—Consider now the case when the gas, at a temperature T say, is being traversed by radiation from an external source at a much higher temperature T_1, e.g. from the sun considered as a black body at 6500°K. In this case the gas and the radiation traversing it are not in thermodynamic equilibrium and the number of molecules dissociated per second is not controlled by the temperature of the gas but by the temperature of the external radiating body whose ultra-violet radiation will cause photo-dissociation.

If $I_\nu d\nu$ be the amount of energy within the frequency range ν and $\nu+d\nu$ passing through unit area of the gas per second, then the number of absorption processes per second per unit volume is

$$n_{AB}\psi_\nu I_\nu d\nu/h\nu \,. \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \qquad (7a)$$

The corresponding number of spontaneous and stimulated recombinations is

$$4\pi n_{A'} n_B \beta_\nu \left(\frac{M}{2\pi kT}\right)^{\frac{3}{2}} \left(1+\frac{c^2 I_\nu}{8\pi h\nu^3}\right) e^{-\frac{MV^2}{2kT}} V^3\, dV. \qquad \ldots \qquad (10a)$$

It should be noted that in this case expressions (7*a*) and (10*a*) will not be equal for every frequency as in the previous case of thermal equilibrium but the total energy involved in all the absorption processes taken together must be equal to that for all the recombinations. This is fulfilled by multiplying both (7*a*) and (10*a*) by $h\nu$, integrating and equating as follows:

$$\int_{\nu_0}^{\infty} n_{AB}\psi_\nu I_\nu d\nu = \int_{V=0}^{V=\infty} 4\pi n_{A'} n_B \beta_\nu \left(\frac{M}{2\pi kT}\right)^{\frac{3}{2}} \left(1+\frac{c^2 I_\nu}{8\pi h\nu^3}\right) e^{-\frac{MV^2}{2kT}} V^3 h\nu\, dV.$$

Substituting for β_ν and converting V in terms of ν by (8)

$$\frac{n_{A'} n_B}{n_{AB}} = \frac{c^2}{8\pi h}\left[\frac{n_{A'} n_B}{n_{AB}}\right]_0 \frac{\int_{\nu_0}^{\infty} \psi_\nu I_\nu\, d\nu}{\int_{\nu_0}^{\infty} \psi_\nu \nu^3 \left(1+\frac{c^2 I_\nu}{8\pi h\nu^3}\right) e^{-h\nu/kT}\, d\nu} \qquad (13)$$

For the case under consideration the atmosphere is receiving diffuse dilute radiation within a small solid angle Ω from the black body (sun at 6500°K.) at a large distance. The value of I_ν at the top of the atmosphere is given by

$$I_\nu = W\,\frac{8\pi h\nu^3}{c^2} \cdot \frac{1}{e^{-h\nu/kT_1}-1},$$

where W—the dilution factor $= \dfrac{\Omega}{4\pi} = \dfrac{R^2}{4r^2}$,

R—radius of the sun $= 6{\cdot}95\times10^{10}$ cms.,

r—sun's distance from earth $= 1{\cdot}494\times10^{13}$ cms.

The intensity is further weakened as the radiation penetrates into the atmosphere. If N be the total number of absorbing gas molecules above one sq. cm. at the desired level having absorption coefficient ψ_ν, the reduced intensity at that level is

$$I_\nu = W\frac{8\pi h\nu^3}{c^3}\cdot\frac{1}{e^{h\nu/kT_1}-1}e^{-\psi_\nu N}.$$

Thus finally

$$\frac{n_{A'}n_B}{n_{AB}} = \left[\frac{n_{A'}n_B}{n_{AB}}\right]_0 W\frac{\displaystyle\int_{\nu_0}^{\infty}\frac{\psi_\nu\nu^3}{e^{h\nu/kT_1}-1}e^{-\psi_\nu N}d\nu}{\displaystyle\int_{\nu_0}^{\infty}\psi_\nu\nu^3\left(1+\frac{We^{-\psi_\nu N}}{e^{h\nu/kT_1}-1}\right)e^{-h\nu/kT}d\nu} \quad \text{.. (14)}$$

If x denotes the degree of dissociation, i.e. the fraction of molecules AB dissociated into atoms A' and B then for N molecules/cm.^{3} *before* dissociation, we get

$N(1-x)$ molecules AB,
Nx atoms A',
and Nx atoms B,

i.e. a total of $N(1+x)$ particles/cm.^{3} *after* dissociation. Therefore if p_d represents the equilibrium pressure due jointly to AB, A' and B after dissociation, then

$$p_d = p(1+x),$$

where p = pressure due to AB *before* dissociation, expressed in dynes/cm.^{2} Therefore

$$\frac{n_{A'}n_B}{n_{AB}} = \frac{Nx^2}{1-x} = \frac{x^2}{1-x}\cdot\frac{p}{kT},$$

i.e.,

$$\frac{x^2}{1-x} = \frac{kT}{p}\cdot\frac{n_{A'}n_B}{n_{AB}}$$

$$= \frac{kT}{p}\left[\frac{n_{A'}n_B}{n_{AB}}\right]_0 W\frac{\displaystyle\int_{\nu_0}^{\infty}\frac{\psi_\nu\nu^3}{e^{h\nu/kT_1}-1}e^{-\psi_\nu N}d\nu}{\displaystyle\int_{\nu_0}^{\infty}\psi_\nu\nu^3\left(1+\frac{We^{-\psi_\nu N}}{e^{h\nu/kT_1}-1}\right)e^{-h\nu/kT}d\nu} \quad \text{.. (15)}$$

The integral in the denominator on the right hand side is practically

$$= \int_{\nu_0}^{\infty}\psi_\nu\nu^3 e^{-h\nu/kT}d\nu$$

on account of the extreme smallness of W. To evaluate this we must know the values of absorption coefficient ψ_ν for all frequencies within the

absorption band. Our only knowledge of this in case of oxygen is from the experimental results of Ladenburg and Van Voorhis [11] and the values of ψ_ν in Table I are obtained from their paper. Further, considering the rapid variation of $e^{-h\nu/kT}$ with ν as seen from Table I, the quantity $\psi_\nu \nu^3$ may, for the purpose of integration, be regarded as a constant having the value corresponding to ν_0 for which the value of $e^{-h\nu/kT}$ is a maximum.

TABLE I

$T = 300°$K (average value for the region concerned)

$\nu \times 10^{-15}$	$e^{-h\nu/kT}$	$\psi_\nu \times 10^{17}$	$\psi_\nu \nu^3 \times 10^{-28}$
1·71	$1{\cdot}26 \times 10^{-117}$	0·13	0·67
1·75	$4{\cdot}51 \times 10^{-121}$	0·22	1·18
1·80	$1{\cdot}97 \times 10^{-126}$	0·37	2·16
1·90	$2{\cdot}07 \times 10^{-131}$	0·80	5·49

$$\text{Hence} \quad \int_{\nu_0}^{\infty} \psi_\nu \nu^3 e^{-h\nu/kT} d\nu \doteqdot [\psi_\nu \nu^3]_{\nu=\nu_0} \times \int_{\nu_0}^{\infty} e^{-h\nu/kT} d\nu$$

$$= [\psi_\nu \nu^3]_{\nu=\nu_0} \times \frac{kT}{h} e^{-h\nu_0/kT} = 6{\cdot}702 \times 10^{27} . \frac{kT}{h} e^{-h\nu_0/kT} \quad .. \quad (16)$$

Substituting in Eq. (15) and introducing the numerical values of the quantities involved, we finally get

$$\frac{x^2}{1-x} = 8{\cdot}588 \times 10^{-37} . \frac{\sqrt{T}}{p} \int_{\nu_0}^{\infty} \psi_\nu \nu^3 \frac{e^{-\psi_\nu H}}{e^{h\nu/kT_1}-1} d\nu.$$

In this equation p the partial pressure of O_2 at the level considered is expressed in dynes/cm.^{2} Expressing p in mm. of mercury, the formula reduces to

$$\frac{x^2}{1-x} = \frac{6{\cdot}527 \times 10^{-40} \sqrt{T}}{p} \int^{\infty} \psi_\nu \nu^3 \frac{e^{-\psi_\nu H}}{e^{h\nu/kT_1}-1} d\nu. \quad .. \quad (17)$$

In order to find the value of x at different levels a knowledge of temperature and partial pressure of O_2 at each level is necessary. These data are not known with certainty but those that are considered as most plausible may be used.

Regarding temperature we may assume, after Martyn and Pulley [12], a value of 160°K. at 80 km. and a linear temperature gradient, increasing with height at the rate of 4°K. per km. This is on the assumption that the temperature in the vicinity of E_1-layer is about 250°K.

Regarding partial pressure we may assume that if O_2 were not dissociated, the condition of thorough mixing of N_2 and O_2 would have been maintained up to at least 130 km. as exists at lower levels. This is justified because even up to this height the tendency for diffusive separation is

negligibly small [13]. The pressure in this region as obtained from radio observations is 10^{-3} mm. at 100 km. Since this corresponds to about 0·2 cm. of O_2 at S.T.P. most of the dissociating radiation will be absorbed higher up. We can therefore assume that the atmospheric composition at 95 km. is not far different from that at the ground and that N_2 and O_2 exist almost in the same proportion, namely, 20 per cent O_2 and 80 per cent N_2 in round figures. Starting with this pressure and assuming the temperature distribution as indicated above, the partial pressure of O_2 at different heights (if there were no dissociation) can be calculated.

The initial total pressure P_z due to N_2 and O_2 at any level z cm. above 80 km. (in the 80–130 km. region) is thus given by

$$P_z = P_{80}\,(1+\alpha z)^{-\frac{gm}{\alpha T_0 k}},$$

where α is the coefficient of increase of temperature with height above 80 km. level ($= 2{\cdot}5\times10^{-7}$°K. per cm. per degree); T_0 is the temperature at 80 km. level ($=160$°K.); g is the acceleration due to gravity; m is the mean molecular mass of the air in the region (in gm.) and k ($=1{\cdot}371\times10^{-16}$ erg per degree), the universal gas constant. On substitution,

$$P_z = P_{80}\,(1+\alpha z)^{-8\cdot5}.$$

Assuming $P_{100} = 10^{-3}$ mm., we find $P_{80} = 3{\cdot}14\times10^{-2}$ mm. and hence $P_z = 3{\cdot}14\times10^{-2}\,(1+\alpha z)^{-8\cdot5}$ mm. Table II gives the values of P_z and p_z, the partial pressure due to O_2, at different heights above 80 km.

TABLE II

z (km.)	T	P_z (mm.)	p_z (mm.)	z (km.)	T	P_z (mm.)	p_z (mm.)
5	180	$1{\cdot}15\times10^{-3}$	$2{\cdot}30\times10^{-3}$	30	280	$2{\cdot}70\times10^{-4}$	$5{\cdot}40\times10^{-5}$
10	200	$4{\cdot}71\times10^{-3}$	$9{\cdot}42\times10^{-4}$	35	300	$1{\cdot}50\times10^{-4}$	$3{\cdot}00\times10^{-5}$
15	220	$2{\cdot}09\times10^{-3}$	$4{\cdot}18\times10^{-4}$	40	320	$8{\cdot}68\times10^{-5}$	$1{\cdot}74\times10^{-5}$
20	240	$1{\cdot}00\times10^{-3}$	$2{\cdot}00\times10^{-4}$	45	340	$5{\cdot}19\times10^{-5}$	$1{\cdot}04\times10^{-5}$
25	260	$5{\cdot}05\times10^{-4}$	$1{\cdot}01\times10^{-4}$	50	360	$3{\cdot}19\times10^{-5}$	$6{\cdot}38\times10^{-6}$

(iii) *Dissociation at different levels.*—The quantity $\mathbf{N}$ in Eq. (17) is given by

$$\mathbf{N} = pHN,$$

where p is partial pressure of O_2 before dissociation expressed in atmospheres, H is height of the homogeneous atmosphere for O_2 and N is number of molecules of gas per cm.^{3} at S.T.P. Hence $\mathbf{N} = 2{\cdot}861\times10^{22}p$, where p is expressed in mm. of mercury. Eq. (17) thus finally reduces to

$$\frac{x^2}{1-x} = \frac{6{\cdot}527\times10^{-40}\sqrt{T}}{p}\int_{\nu_0}^{\infty}\psi_\nu\nu^3\,\frac{e^{-2\cdot861\times p\psi_\nu 10^{22}}}{e^{h\nu/kT_1}-1}\,d\nu. \quad .. \quad (18)$$

The values p and T for any level are obtained from Table II and the integral is evaluated by graphical method. Table III gives the values of the fraction x of O_2-molecules dissociated at various heights, the equilibrium concentrations n_{O_2} of O_2 and n_O of O and also the total amounts of O_2 and O above each level in the region studied.

TABLE III

Height above ground (km.)	Degree of dissociation x	n_{O_2}	n_O	Amount of gas (cm. at S.T.P.)	
				O_2	O
85	0·00	$1·243\times10^{14}$	0·00	2·302	0·260
90	$2·96\times10^{-8}$	$4·581\times10^{13}$	$2·712\times10^{6}$	0·866	0·260
95	$7·10\times10^{-4}$	$1·847\times10^{13}$	$2·624\times10^{10}$	0·313	0·258
100	$6·70\times10^{-2}$	$7·561\times10^{12}$	$1·086\times10^{12}$	$8·570\times10^{-2}$	0·251
105	0·630	$1·398\times10^{12}$	$4·761\times10^{12}$	$1·177\times10^{-2}$	0·190
110	0·956	$8·250\times10^{10}$	$3·585\times10^{12}$	$5·885\times10^{-4}$	0·113
115	0·994	$5·835\times10^{9}$	$1·752\times10^{12}$	$5·254\times10^{-5}$	$6·33\times10^{-2}$
120	0·998	$5·288\times10^{8}$	$1·057\times10^{12}$	$6·336\times10^{-6}$	$3·68\times10^{-2}$
125	0·999	$8·925\times10^{7}$	$5·95\times10^{11}$	$1·383\times10^{-6}$	$2·20\times10^{-2}$
130	0·999	$1·724\times10^{7}$	$3·45\times10^{11}$	$4·000\times10^{-7}$	$1·34\times10^{-2}$

It will be seen from Fig. 1 that the density of O_2-molecules decreases very rapidly with height above 100 km. At the same time the density of the oxygen atoms which is almost zero at 80 km. increases rapidly with height, attains a maximum at about 105 km. and then gradually decreases. The transition layer in which the density of O_2 rapidly decreases plays an extremely important part in the production of the *E*-layer of the ionosphere. This is discussed in Chapter VI.

(*c*) Results of other investigations

Dissociation of O_2 in the upper atmosphere has also been studied by several other workers. We describe below briefly the results obtained.

Chapman [14] was the first to consider the dissociation of O_2-molecules and recombination of O-atoms in the upper atmosphere and estimated the position of the region of transition $O_2 \rightarrow O+O$.

According to Wulf and Deming's [3] calculation the concentration of O_2-molecules is less than 2 per cent at 100 km. and the height of half dissociation of the same is 89 km. The shape of the transition layer is similar to that given in Fig. 1. Chapman, and also Wulf and Deming, assume that the O-atoms recombine to form O_2-molecules by three-body collision process.

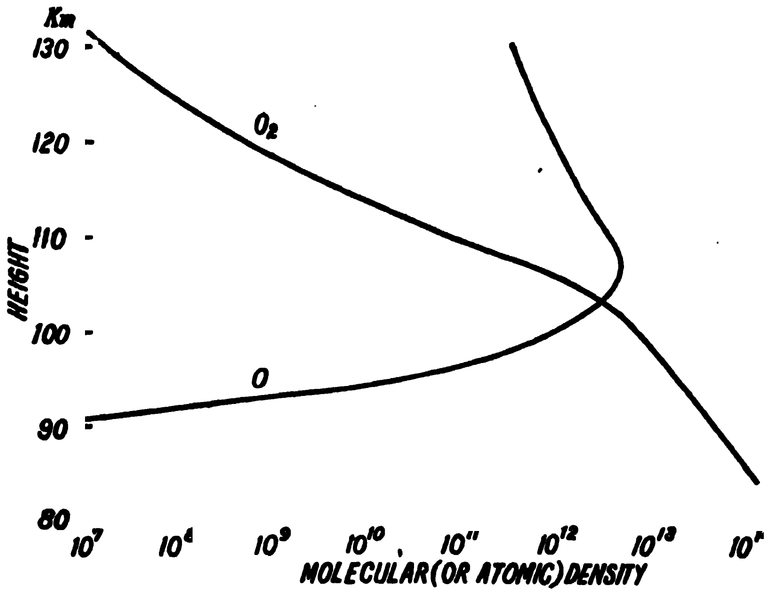


FIG. 1. Distribution of atomic and molecular oxygen in the upper atmosphere. Due to dissociative action of solar ultraviolet radiation ($\lambda < 1751$) the density of molecular oxygen rapidly diminishes above 95 km. Above 130 km. the atmosphere consists practically of N_2 and O. (After Rakshit.)

Penndorf [4] has calculated the vertical distribution of oxygen atoms using different values of intensity of the incoming radiation (corresponding to different assumed temperatures of the sun), different zenith distances of the sun, and different rate constants (coefficient of recombination by three-body collisions). It is found that the thickness of the transition layer is about 10 km. and its mean height lies between 100 and 110 km. It is also shown that the effects of varying the intensity of the incoming radiation and of the rate constant are large, but the effect of variation of the zenith distance of the sun is of little importance.

3. ATOMIC NITROGEN IN THE UPPER ATMOSPHERE

Dissociation process.—As already mentioned there is no simple photodissociation process for N_2 analogous to that given by Eq. (1). Of the various probable processes suggested [15], dissociative recombination of N_2^+ ions seems most promising: There is evidence of the presence of N_2^+ ions in the upper atmosphere both in the auroral regions and also in other parts where solar radiation is acting. (See Chap. VIII, Sec. 6 and Chap. X, Sec. 4.) These ions, on collision with electrons, undergo dissociative recombination, that is, the ion on being neutralized dissociates into (excited) atoms. According to Bates [16] such dissociative recombination process has a high probability, the coefficient being of the order of 10^{-7} cm.^{3}/s. Mitra has

discussed this process and has shown that, under certain condition, the following reaction has a high probability [17],

$$N_2^+(X') + e \rightarrow N(^2P) + N(^2D) \quad \ldots \quad \ldots \quad \ldots \quad (19)$$

Here, the energy supplied by the left-hand side (if the N_2^+ ion is in the lowest vibrational level $v''=0$) is 15·58 eV, the ionization energy of N_2. The energy demanded by the right-hand side is the total of the dissociation energy of the N_2-molecule (9·76 eV, Gaydon's value [18]) and the energies of the two excited N-atoms (3·56 eV+2·37 eV = 5·93 eV), i.e. 15·96 eV. The demand is thus in excess of supply by 0·11 eV. This deficiency in supply is made up if the reacting $N_2^+(X')$ ion be in some low vibrational level $v'' > 0$, instead of in the lowest level. The presence of $N_2^+(X')$ ions in such vibrational levels is evidenced by the spectral composition of the first negative bands. The reaction also explains the emission of the observed atomic nitrogen lines in aurorae (see Chap. VIII, Sec. 3(*a*)).

Abundance of nitrogen atoms.—Various estimates of the proportion of nitrogen atoms in the regions of auroral display have been made. According to Dufay [19], at very great auroral heights, the nitrogen atoms in the ^2D state (metastable life 8 hours!) are seven times as numerous as the oxygen atoms in the ^1D state (from which the red auroral lines are emitted). According to Gauzit [20] the nitrogen (like oxygen) is mostly in the atomic state above 100 km. and the proportion of N_2 falls rapidly to 10^{-4} at greater heights. This latter conclusion, however, seems hardly justified, because, the emissions of N_2 and N_2^+ bands (specially of the latter, the first negative bands) from the highest limits of aurorae indicate that molecular nitrogen is present in appreciable proportion in such regions. The estimate of Bernard [21], namely, that the dissociation of N_2 is one-fifth at the base of the aurora and three-fourths at the upper limit, seems more reasonable. (See also ref. [22].)

It is to be noted that the presence of any atomic nitrogen lines in the night sky air-glow spectrum is very doubtful. This, however, does not mean that nitrogen in the atomic state is not present in the quiet night sky. It merely shows that there may not be present in the quiet night sky any effective mechanism for excitation of the N-atoms. It is to be mentioned that N_2^+ ions are produced by absorption of solar radiation $\lambda < 795$A. Hence the dissociative recombination of N_2^+ ions, as indicated by Eq. (19), must be going on in all parts of the globe.

The contemporary state of knowledge regarding the prevalence of atomic nitrogen in the upper atmosphere may perhaps be thus summarized: Nitrogen in the atomic state *is* present in the upper atmosphere specially in the auroral regions. Its proportion in regions removed from auroral displays is also not inconsiderable. It is, however, not so abundantly present as atomic oxygen.

CHAPTER VI

THE IONOSPHERE

1. INTRODUCTION—HISTORICAL

Perhaps the most important effect of solar ultraviolet radiation on the upper atmosphere is the ionization of its constituent gases. The ionization commences from a height of about 60 km. and extends up to the highest limits of the atmosphere. The density of ionization is not uniform throughout; there are 'regions' or 'layers' of maximum ionization. These ionized regions—known collectively as the *ionosphere**—play a fundamental rôle

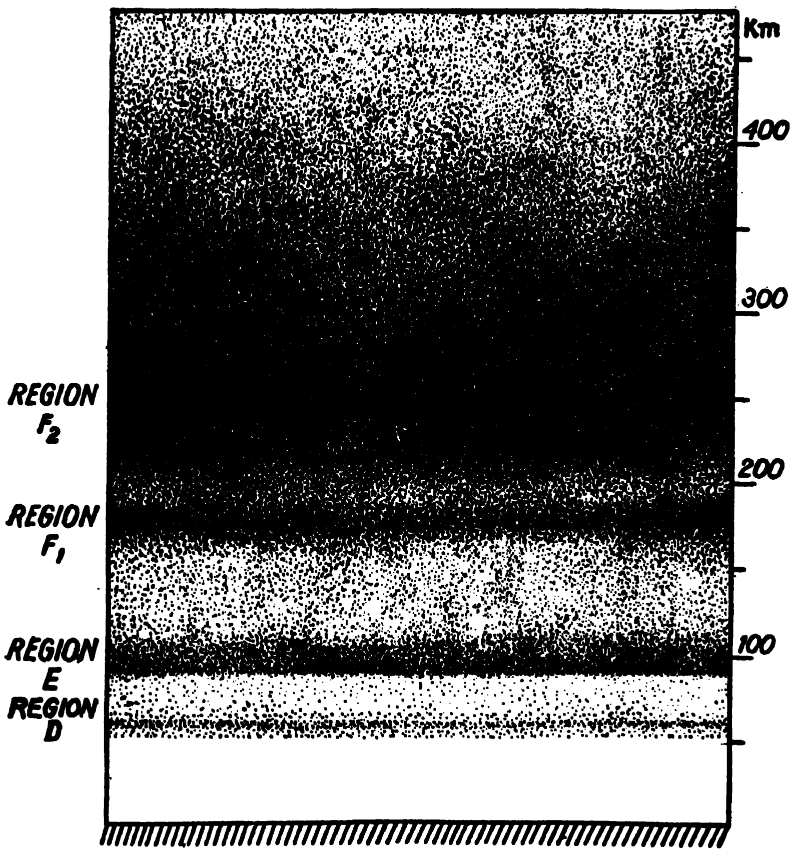


FIG. 1. Ionized regions of the upper atmosphere. The relative density of ionization is roughly indicated by the depth of shading. During daytime another region of ionization E_2 is sometimes formed immediately above E; this is not shown. Région D is mainly an absorbing region. At night this region disappears and also F_1 and F_2 coalesce to form a single Region F.

* The term *ionosphere* is due to Watson Watt.

in the propagation of radio waves round the curved surface of the earth; but for their presence, long distance radio communication would have been an impossibility. Fig. 1 depicts roughly the region of the atmosphere occupied by the ionospheric 'layers'.

The success of Marconi in 1901 in sending wireless signals from Cornwall to Newfoundland was the source of considerable discussion in scientific circles as to the mode of propagation of wireless waves round the curved surface of the Atlantic. Since wireless waves are nothing but electromagnetic waves as constitute radiant heat and light, it was natural to suppose that the phenomenon was one of diffraction and physicists and mathematicians like Macdonald [1], Rayleigh [2], Poincaré [3] and others [4, 5, 6] directed their attention towards solving the problem of the diffraction of the waves by a spherical surface like that of the earth. Calculations showed, however, that the diffraction effect was quite inadequate to explain the observed bending of the waves round the curved surface of the earth.

In 1902, Kennelly [7] in America and Heaviside [8] in England almost simultaneously postulated the existence of a conducting layer in the upper regions of the earth's atmosphere and suggested that this layer might deflect the radio waves and force them to follow the curvature of the earth. Heaviside further made the suggestion that the conductivity of this region is due to the presence of positive and negative ions produced, most probably, by the ionizing action of solar radiation. It should, however, be mentioned here that as far back as 1878, Balfour Stewart [9] postulated the existence of conducting strata in the upper atmosphere in connection with his theory of the daily variation of the earth's magnetic field (see Chapter VII).

The actual manner in which the charged particles affect the propagation of radio waves was first pointed out by Eccles [10] in 1912. Some essential details which were lacking in Eccles' theory and for which it lay undeveloped were supplied by Larmor [11] in 1924. The Eccles-Larmor theory is now regarded as the basic theory of radio wave propagation in the ionosphere and has been supplemented by the so-called magneto-ionic theory developed by Appleton, Hartree, Goldstein and others.

The conducting upper atmosphere, though postulated by Kennelly and Heaviside in 1902, lacked direct evidence as to its existence till 1924. Agreement of quantitative measurements on long-wave field strength with Watson-Eckersley formula, and the many peculiar features of short-wave propagation discovered since 1921, pointed towards the existence of such a layer. But these phenomena were only evidence of an indirect nature. The first direct evidence came in 1925 when Appleton and Barnett [12], by comparing the intensities of fading of signals received simultaneously on a loop and on a vertical aerial, proved the existence of indirect or sky waves coming down after reflection from the Kennelly-Heaviside layer. Immediately afterwards, Smith-Rose and Barfield [13] detected the presence of downcoming waves by employing direction-finding apparatus of their own design. These experiments not only gave direct evidence of the reflecting

layer but also disclosed some of its properties, e.g. its height, reflecting power. But perhaps the most striking evidence was that furnished by experiments by Breit and Tuve [14]. These authors showed that a wave packet or a 'pulse' of small duration sent out from a transmitter produced two (sometimes more) impulses instead of one in a receiver placed a few kilometres from the transmitter. The obvious conclusion was that the impulse coming first was due to the direct wave travelling along the ground and those coming later were due to indirect waves or echoes reflected from the ionosphere. This method—the so-called 'pulse' or group-retardation method—developed and improved by various workers is now the most powerful method of investigating the ionosphere.

The radio wave is thus one of the most useful tools at our disposal for studying the high regions of the atmosphere hundreds of kilometres above the surface of the earth. The famous Piccard balloon ascended up to 16 km.; the sounding balloons sent by meteorologists seldom reach beyond 35 km.; the rockets may reach up to 200 km. But, the radio waves can ascend far beyond these heights and, when they come down 'reflected' by the ionospheric layers, they convey with them indelible messages regarding the physical conditions of the upper atmosphere which are quite as useful as those recorded by instruments carried by sounding balloons or rockets. For interpreting these messages it is, however, necessary to make a close study of the mode of propagation of the radio waves through the ionosphere. This has to be made not only for the simple case of propagation in a medium containing ions and electrons, but also for the more complicated case when the ionized medium is traversed by a magnetic field as that due to the earth.

2. PROPAGATION OF ELECTROMAGNETIC WAVES IN AN IONIZED ATMOSPHERE (RAY TREATMENT)

In this section we shall first discuss after Eccles [10] and Larmor [11] the simple case of propagation of electromagnetic waves through a medium containing ions and electrons in the absence of any external magnetic field. We shall then proceed with the more complicated case of propagation in the presence of a magnetic field.

Throughout this section it will be assumed that the extent of any inhomogeneity in the ionized atmosphere is large compared to the wavelength. For example, if there is a horizontally stratified ionized layer the ionization density increasing with height, then we assume that the ionization density does not change very appreciably within the distance of one wavelength. Under such conditions the propagation of a wave may be treated from the point of view of the propagation of a ray after the manner of geometrical optics. This we call 'ray treatment'. The case where ray treatment fails—when the wavelength is large compared to the extent of the inhomogeneity—will be discussed in Sec. 3.

(a) Propagation in the absence of magnetic field—Theory of Eccles and Larmor

(i) *The fundamental equation.*—Let the ionized medium contain N_e electrons per cm.^{3}, each of charge e and mass m. When an electromagnetic wave travels through such a medium, the motion of the electrons is influenced by the electric force of the passing wave. For simplicity, let us suppose the wave front to be plane and the wave plane-polarized. The electrons will then find themselves in an electric field of fluctuating intensity given by $E=E_0 \sin pt$, E_0 being the amplitude of E and p the angular frequency of the wave. The electrons will vibrate under the action of this force and if, for the moment, we neglect the collision of the vibrating electrons with the neighbouring neutral atoms and molecules, we can write the equation of motion of an electron as

$$m\frac{d\xi}{dt} = E_0 e \sin pt, \qquad \dots \quad \dots \quad \dots \quad (1)$$

where ξ is the instantaneous velocity of the electron in the direction of the electric field. Integrating we have

$$\xi = -\frac{E_0 e}{mp}\cos pt, \qquad \dots \quad \dots \quad \dots \quad (1.1)$$

where the integration constant has been made equal to zero by a proper choice of the phase. This motion of the electrons produces a convection current, the density of which is given by

$$N_e e\xi = -\frac{E_0 N_e e^2}{mp}\cos pt, \qquad \dots \quad \dots \quad \dots \quad (1.2)$$

where N_e is the number of electrons per cm.^{3}

Now, at the point under consideration, there is also a displacement current due to the fluctuation of the field intensity given by

$$\frac{K}{4\pi}\frac{\partial E}{\partial t} = \frac{KE_0}{4\pi}p\cos pt, \qquad \dots \quad \dots \quad \dots \quad (2)$$

K being the dielectric constant of the medium itself without the electrons. The total current is, therefore, given by

$$\left[K - \frac{4\pi N_e e^2}{mp^2}\right]\frac{E_0 p}{4\pi}\cos pt.$$

The presence of the electrons has thus the effect of reducing the dielectric constant of the medium from K to $K-4\pi N_e e^2/mp^2$. In empty space, $K=1$, and the phase velocity of electromagnetic waves is equal to c, that of light. The presence of electrons in an otherwise empty space thus reduces the dielectric constant from unity to $1-4\pi N_e e^2/mp^2$, and alters the phase velocity from c to u given by

$$u = \frac{c}{\sqrt{1-\dfrac{4\pi N_e e^2}{mp^2}}}. \qquad \dots \quad \dots \quad \dots \quad (3)$$

The most important effect of the reduction of dielectric constant to a value less than unity is that electromagnetic waves sent upwards from the surface of the earth may be bent down by 'reflection' from the ionized regions of the upper atmosphere. If we imagine that the electron density is increasing from bottom upwards, then it is easy to see that the direction of propagation of an electromagnetic wave incident obliquely on the ionized stratum will be gradually bent down. In Fig. 2, AB represents a small section of the wave front, incident on the ionized stratum represented by the shaded region. Since the top portion of the wave front is moving in a region of greater electronic density, its phase velocity is also greater. The wave front thus veers round and ultimately proceeds downward. It is obvious that the presence of such an ionized region surrounding the earth

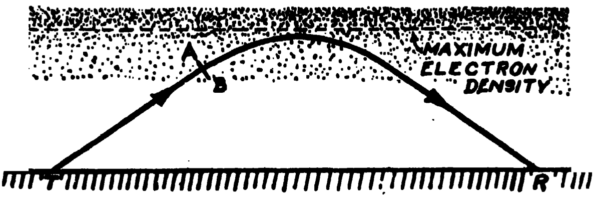


FIG. 2. Bending of the path of a ray by an ionized region. The top of the wave front (AB) being in a region of higher density moves faster. T—Transmitter; R—Receiver.

will have the effect of confining to the earth radio waves which would otherwise escape.

We can easily deduce the condition for which the wave will be totally reflected from the ionized region. We first note from Eq. (3) that the refractive index of the ionized medium is given by

$$\frac{c}{u} = \sqrt{1 - \frac{4\pi N_e e^2}{mp^2}}. \tag{4}$$

Considering Fig. 3, let AB be the lower boundary of the ionized stratum,

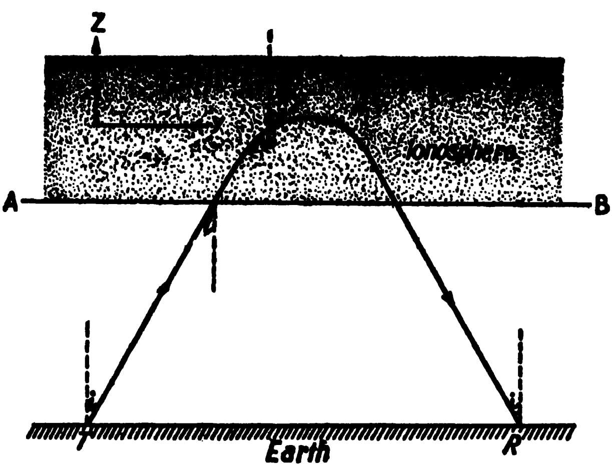


FIG. 3. Path of a refracted ray in the ionosphere.

the ionization increasing from bottom upwards. Let the wave be incident on AB at an angle i to the vertical; then, assuming that the strata of equal ionization are horizontal, and that the refractive index of the region just below AB is unity, the angle ψ which the refracted ray would make with the vertical at any point S is given, according to Snell's Law, by

$$\sin\psi = \frac{\sin i}{\mu_\psi}, \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (5)$$

where μ_ψ is the refractive index of the ionized medium at the point. It is obvious that the direction of the ray will be horizontal, that is, the wave will be totally reflected at the point where $\psi = \pi/2$. If μ_0 represents the refractive index of the medium at this point, we get for the condition of total reflection,

$$\frac{\sin i}{\mu_0} = 1$$

$$\text{or} \qquad \sin i = \mu_0 = \sqrt{1 - \frac{4\pi N_e^0 e^2}{mp^2}}, \quad \ldots \quad \ldots \quad (6)$$

where N_e^0 is the value of the electron density at the point of reflection. If the wave is incident on the ionized layer vertically, then $i = 0$, and the condition for total reflection is given by

$$\mu_0 = 0 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (7)$$

$$\text{or} \qquad \frac{4\pi N_e^0 e^2}{mp^2} = 1. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (7.1)$$

A more general condition of wave reflection may be deduced as follows [31]: Let a wave of wavelength λ (*in vacuo*) be incident on a horizontally stratified ionized region at an angle i with the vertical. Let the plane of incidence be the YZ plane, the Y- and Z-axis being along the horizontal and vertical directions respectively (Fig. 3).

At the point (y, z) in the path of the ray, let ψ be the angle between the direction of propagation of the wave and the vertical. The wave function appropriate to this level may then be written in the form

$$\exp\left[j \cdot \frac{2\pi}{\lambda}\{ct - \mu_\psi(\sin\psi \cdot y + \cos\psi \cdot z)\}\right], \quad \ldots \quad (8)$$

where μ_ψ is the refractive index of the medium at the level of (y, z). If u_ψ be the velocity of phase propagation at the same level in the direction ψ, then $\mu_\psi = c/u_\psi$.

Now imagine μ_ψ as a vector, the direction of which is that of the propagation of the wave. The horizontal component of this vector is then $\mu_\psi \sin\psi$ (Fig. 4). By Snell's Law this is equal to $\sin i$, and is therefore a constant. The vertical component is $\mu_\psi \cos\psi$ and is variable. Putting q for $\mu_\psi \cos\psi$, the above wave function can be written as

$$\exp\left[j \cdot \frac{2\pi}{\lambda}\{ct - \sin i \cdot y - qz\}\right]. \quad \ldots \quad \ldots \quad (9)$$

The propagation of the wave through the ionized layer can now be completely described by this wave function if the variation of q with z in the medium is known.

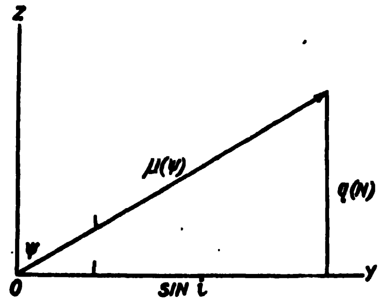


FIG. 4. Illustrating phase propagation.

The variation of q with z in a horizontally stratified ionized region in which N_e increases with height is depicted in Fig. 5. At the point of incidence, represented by I in the figure $q = \cos i$, for $\psi = i$ and $\mu_\psi = 1$.

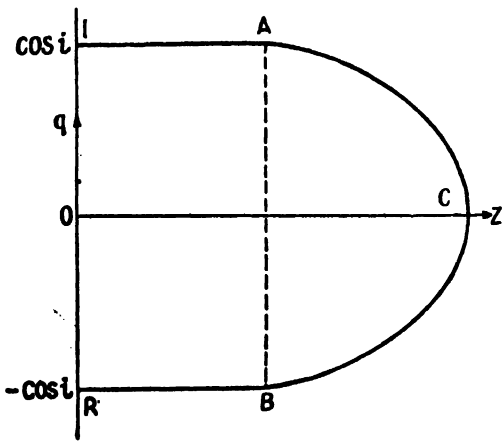


FIG. 5. Variation of q (vertical component of μ_ψ regarded as a vector) with height.

As the wave proceeds upwards, both μ_ψ and $\cos\psi$ decrease, so that q gradually diminishes until it reaches zero value at the level where ψ is equal to $\pi/2$, represented by the point C in the figure. As the value of ψ increases further, the curve turns back towards the q-axis reaching it at the point R where $q = -\cos i$, being symmetrical with respect to the top curve. The physical significance of this curve is obvious. At the point C, $\psi = \pi/2$ and this is the point where the upgoing wave is reflected down. In fact, the top half of the curve may be taken as representing the propagation of the upgoing wave and the lower half that of the downcoming wave. At the turning point C,

$$\frac{dq}{dz} = \frac{d}{dz}(\mu_\psi \cos\psi) = \infty . \quad .. \quad .. \quad .. \quad (10)$$

This condition, namely, that the wave is totally reflected at the point where $dq/dz = \infty$, is more general than the condition $\mu = 0$ for vertical propagation or $\mu = \sin i$ for oblique propagation. It will be noted that for vertical propagation, q is identical with μ and Eq. (10) reduces to

$$\frac{d\mu_\psi}{dz} = \infty. \qquad (11)$$

As mentioned in the introductory paragraph, the conditions deduced above are based on the ray theory of wave propagation. This theory obviously breaks down when $\mu \to 0$ because at this point the wavelength λ in the medium tends to infinity. Nevertheless, though we cannot say what exactly is the mode of propagation at the point of reflection, the simple ray treatment gives us the trajectory of the ray correctly immediately before it enters into and after it emerges out of the region of reflection.

It has so far been assumed that the charged particles are electrons. The same treatment would hold good for ions, the only difference being that ions being heavier than electrons, a much larger number density would be necessary to produce the same increase in the phase velocity and consequently to produce the same bending in the path of the electromagnetic wave. It was Larmor who first pointed out the relative importance of electrons on account of their low mass in modifying the propagation of electromagnetic waves in an ionized medium.

(ii) *Effect of collision.*—In writing out the fundamental equation (1), we neglected collisions, if any, of the vibrating electrons with neighbouring particles. This we now take into account, because in the case of the ionosphere, collisions always occur due to the presence, in large numbers, of neutral particles and heavy ions besides electrons. The collisions destroy in part the directed momentum of the vibrating electrons, the ultimate effect of which is to cause absorption of energy from the electromagnetic waves traversing the ionized medium. The ionosphere, in fact, behaves as an absorbing medium of conductivity *

$$\sigma = \frac{N_e e^2 \nu}{m(p^2+\nu^2)},$$

where ν is the frequency of collisions per second. As such, it has a complex dielectric constant κ' given by

$$\kappa' = \kappa - \frac{4\pi j\sigma}{p}, \qquad (12)$$

and a complex refractive index M given by

$$M^2 = \kappa' = \left(\mu - \frac{jck}{p}\right)^2, \qquad (13)$$

where μ is the refractive index of the medium and k the absorption per unit length of path so that ck/p is the absorption per length of path $\lambda/2\pi$, λ being the wavelength *in vacuo*.

* For the proofs of the equations in this section see Appendix 3.

Further, for the value of σ for the ionized medium given above,

$$M^2 = 1+\frac{1}{\alpha+j\beta}, \qquad \text{(14)}$$

where
$$\alpha = -\frac{mp^2}{4\pi N_e e^2};\ \beta = \frac{mp\nu}{4\pi N_e e^2}.$$

Hence from Eqs. (13) and (14), separating the real and imaginary parts,

$$\mu^2-\left(\frac{ck}{p}\right)^2 = 1+\frac{\alpha}{\alpha^2+\beta^2} = 1-\frac{4\pi N_e e^2}{m(p^2+\nu^2)} \qquad \text{(15)}$$

and

$$2\frac{ck}{p}\mu = \frac{\beta}{\alpha^2+\beta^2} = \frac{4\pi N_e e^2\nu}{mp(p^2+\nu^2)}. \qquad \text{(16)}$$

Or, from Eqs. (15) and (16),

$$2\mu^2 = \left(1+\frac{2\alpha+1}{\alpha^2+\beta^2}\right)^{\frac{1}{2}}+\frac{\alpha}{\alpha^2+\beta^2}+1 \qquad \text{(17)}$$

$$2\left(\frac{ck}{p}\right)^2 = \left(1+\frac{2\alpha+1}{\alpha^2+\beta^2}\right)^{\frac{1}{2}}-\frac{\alpha}{\alpha^2+\beta^2}-1. \qquad \text{(18)}$$

Fig. 6 shows how μ^2 and $(ck/p)^2$ vary with the electron number density N_e

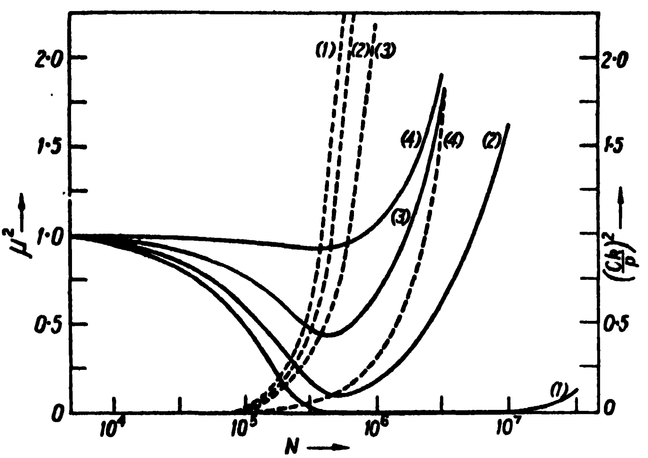


FIG. 6. Illustrating the variations of μ^2 and $(ck/p)^2$ with electron number density N_e for different values of collisional frequency ν and for a wave frequency of 4 Mc/s. The continuous curves are for μ^2 and the broken line curves are for $(ck/p)^2$.
Curves (1) for $\nu = 10^6$; (2) for $\nu = 10^7$; (3) for $\nu = 2{\cdot}51\times10^7$; (4) for $\nu = 10^8$.

for the value of p corresponding to 4 Mc/s. The curves are drawn for the following values of ν: (1) $\nu = 10^6$, i.e., $\nu \ll p$; (2) $\nu = 10^7$, i.e., $\nu < p$; (3) $\nu = p = 2{\cdot}51\times10^7$; and (4) $\nu = 10^8$, i.e., $\nu > p$. The general trend of the curves shows that for any value of ν, the value of μ^2 falls or has a tendency to fall below unity; it passes through a minimum and then rises rapidly with increasing values of N_e.

If k is small, so that $(ck/p)^2$ may be neglected compared with μ^2, the expression for μ^2 assumes the simple form

$$\mu^2 = 1 - \frac{4\pi N_e e^2}{m(p^2+\nu^2)}. \quad .. \quad .. \quad .. (15.1)$$

This expression, however, holds good only over a very limited range. The true value of μ^2 begins to depart considerably from that given by the simple expression, as $\frac{4\pi N_e e^2}{m(p^2+\nu^2)}$ tends to unity. In fact, when the collisional term ν is present, μ^2 according to Eq. (17) cannot become zero or negative, as it could otherwise be if ν were zero. It is to be noted, however, that the expression on the right of Eq. (15.1) is *rigorously* equal to the dielectric constant, because the dielectric constant κ is the real part of κ' and is given by (Eq. (15)),

$$\kappa = \mu^2 - \left(\frac{ck}{p}\right)^2 = 1 - \frac{4\pi N_e e^2}{m(p^2+\nu^2)}.$$

The dielectric constant κ is thus always less than unity even when there is collision though μ^2 (according to (17)) can be greater or less than unity. As is well known, for an absorbing medium, μ^2 cannot be identified with κ. By equating the real and imaginary parts of the right hand sides of (12) and (13) it can be easily shown that

$$\mu^2 = \frac{\kappa}{2} + \sqrt{\frac{\kappa^2}{4} + \frac{4\pi^2\sigma^2}{p^2}}. \quad .. \quad .. \quad .. (19)$$

If, however, the medium be non-conducting so that $\sigma = 0$, then

$$\mu^2 = \kappa.$$

We have in the preceding section deduced that the condition of wave reflection for vertical incidence is $\mu = 0$, or more generally $d\mu/dz = \infty$. Now with collision present the value of the refractive index can neither be equal to zero nor can its rate of change with height $d\mu/dz$ be infinity (Fig. 6). Hence, we cannot have total reflection according to the ray theory. However, in optics we know that there is reflection where there is a sudden change in the refractive index of the medium. The same principle applies in the present case so that in spite of the fact that μ cannot be zero, it can be shown that there will be copious reflection if $d\mu/dz$ is large, i.e., the rate of change of μ per $\lambda/2\pi$ of height (λ in vacuum) is large compared to $2\mu_0$ where μ_0 is the minimum value of μ. If, however, the rate of change of μ per $\lambda/2\pi$ (λ in vacuum) is small compared to $2\mu_0$, then there is only partial reflection. And, if $2\mu_0$ is large compared to the rate of change of μ per $\lambda/2\pi$ (λ in vacuum), there is no reflection, and the wave sent upwards is absorbed. These points will be further discussed in Sec, 3.

(*b*) Propagation in the presence of magnetic field—The magneto-ionic theory—Appleton-Hartree formula

We now discuss the theory of propagation of electromagnetic waves through an ionized medium in the presence of an external magnetic field.

This theory is known as the magneto-ionic theory. Lorentz was the first to treat theoretically the propagation phenomenon in two particular cases, namely, propagation along the magnetic field and propagation at right angles to the same.

Appleton in 1925 [15] first applied the theory to explain the phenomenon of propagation of radio waves through the upper ionized regions of the earth's atmosphere. He pointed out that the quantity He/mc, the gyro-magnetic frequency of the electron about the earth's magnetic field, is comparable with radio frequencies. In 1927 [16] he generalized Lorentz's treatment and derived formulæ for dispersion, absorption and polarization for any angle of inclination of the wave-normal with the magnetic field. At about the same time Goldstein [17] by a different method arrived at the same formula. Later Hartree [18] also derived the formula independently.

If the direction of wave propagation is along the positive direction of the X-axis of a right handed co-ordinate system and the external magnetic field is assumed to be in the XZ plane as shown in Fig. 14, then, according to the magneto-ionic theory, the complex refractive index M and the state of polarization R of the wave are given by

$$M^2 = \left(\mu - \frac{jck}{p}\right)^2$$

$$= 1 + \frac{2}{2(\alpha+j\beta) - \dfrac{\gamma_T^{\,2}}{1+\alpha+j\beta} \pm \sqrt{\left[\dfrac{\gamma_T^{\,4}}{(1+\alpha+j\beta)^2} + 4\gamma_L^{\,2}\right]}} \quad \ldots \quad (20)$$

$$R = \frac{h_z}{h_y} = -\frac{j}{\gamma_L} \cdot \left[\frac{1}{M^2-1} - (\alpha+j\beta)\right] \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (21)$$

where μ—refractve index,

k—index of attenuation,

c—velocity of electromagnetic waves in free space,

N_e—electron density,

H—imposed magnetic field in gauss,

H_L—component of the magnetic field along the direction of propagation,

H_T—component of the magnetic field transverse to the direction of propagation,

p—angular frequency of the wave,

ν—collisional frequency,

$\alpha = -p^2/p_0^{\,2}$,

$\beta = p\nu/p_0^{\,2}$,

$\gamma = pp_H/p_0^{\,2}$,

$\gamma_{L,T} = pp_{L,T}/p_0^{\,2}$,

$p_0^{\,2} = 4\pi N_e e^2/m$,

$p_H = He/mc$,

$p_{L,T} = H_{L,T}e/mc$,

$h_y, h_z = y$- and z-component of the magnetic vector of the wave,

e, m—charge and mass of electron.

Eqs. (20) and (21) constitute the Appleton-Hartree formula.*

The first equation gives us, on the one hand, the nature of the dispersion, i.e., the relation between the frequency of the wave and the refractive index and, on the other, the absorption index of the medium for the two waves into which the incident wave is split up. The second equation gives the nature of the polarization of both the waves.

In the original formula a term l, called the polarization or the Lorentz term, was added to the expression for α. Darwin [19] has, however, shown that under ionospheric conditions the polarization term cannot occur. This point is further discussed in Sec. 4(f).

(c) Simplified Appleton-Hartree formula (neglecting collision)

(i) *Dispersion.*—The formula is undoubtedly complicated; in order to understand its significance it is convenient to first study some simple cases of propagation after Ratcliffe [20]. We therefore commence by neglecting collision, that is when absorption is zero; this will be taken into account later. By putting β and k equal to zero in Eq. (20) it reduces to

$$\mu^2 = 1+\frac{1}{\alpha-\dfrac{\gamma_T^{\,2}}{2(1+\alpha)}\pm\sqrt{\left[\dfrac{\gamma_T^{\,4}}{4(1+\alpha)^2}+\gamma_L^{\,2}\right]}} \quad \cdots \quad \cdots \quad (22)$$

We note from the formula that for a given set of values of the quantities α, γ_T and γ_L, there are two values of μ. Further, for a given value of the magnetic field and for a given direction of propagation of the waves there are three variables, namely, μ, p and N_e. The formula can, therefore, be represented completely by a surface drawn in a three dimensional space depicting the relation between μ, p and N_e. In practice, curves delineating the relations between μ^2 and N_e or μ and N_e (p=constant) are most useful. This is because in actual ionospheric exploration, waves of a given frequency are sent vertically upwards to meet and penetrate into regions of gradually increasing electron number density.

The dispersion curves may be classified into two distinct types according as the angular frequency of the wave is greater or less than the gyrofrequency, i.e., according as p is greater or less than p_H. For each of the two types, the form of the curves gradually changes as the direction of wave propagation is changed from (say) at right angles to the magnetic field to along the magnetic field, i.e., from the so-called transverse to the longitudinal mode of propagation.

We start by drawing the dispersion curves for the simple and extreme cases, namely, the transverse case and the longitudinal case for the two types $p>p_H$ and $p<p_H$ mentioned above. In discussing these and other illustrative curves, we will specially note the points where $\mu^2=0$. As already mentioned, these are the points which give the conditions under which electromagnetic waves incident vertically on an ionized stratum may suffer total reflection.

* For a derivation of the formula, see Appendix 4.

No external magnetic field

We first give, by way of comparison, dispersion curves for the simplest case of all, namely, that of no magnetic field. In this case $\gamma_L = \gamma_T = 0$. Let us put $x = -1/\alpha = p_0^2/p^2$, so that x is proportional to the electron number density N_e. We thus have from Eq. (22)

$$\mu^2 = 1+\frac{1}{\alpha} = 1-x \qquad \ldots \qquad \ldots \qquad \ldots \qquad (23)$$

and the formula reduces to the case deduced by Eccles (Eq. (4)). The

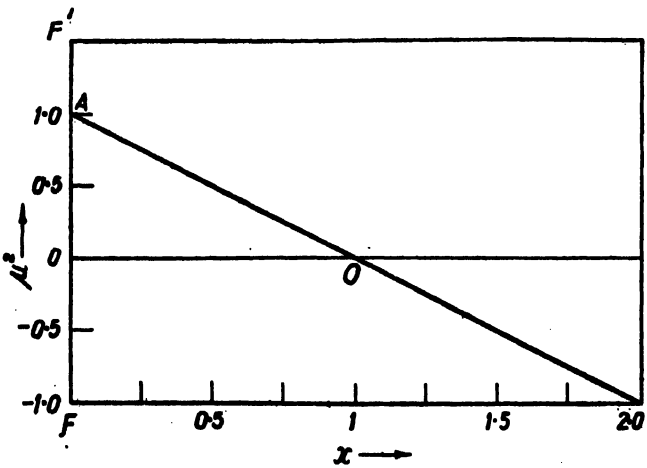


FIG. 7. Variation of μ^2 with x. Case of no magnetic field.

graph connecting μ^2 and x becomes a single straight line (Fig. 7) and the value of $\mu^2 = 0$ where $x = 1$.

Magnetic field along the direction of propagation (longitudinal propagation)

In this case $\gamma_L = \gamma$ and $\gamma_T = 0$. If we put $y = -\gamma/\alpha = p_H/p$, we have from Eq. (22)

$$\mu^2 = 1+\frac{1}{\alpha\pm\gamma} = 1-\frac{x}{1\mp y}. \qquad \ldots \qquad \ldots \qquad \ldots \qquad (24)$$

We first note that unlike the case of no magnetic field, here we have, in general, two values of μ^2 for any value of y, one corresponding to the upper sign in Eq. (24) and the other to the lower. That is, there is magnetic bi-refringence.

Next we note the distinctive behaviours of the dispersion curves in the two cases $p>p_H$ (i.e., $y<1$) and $p<p_H$ (i.e., $y>1$) represented in Fig. 8.

$y<1$

For the upper sign in the equation, the curve swings round from AB to AF' as y increases from 0 to 1. For the lower sign, it swings from AB to AC.

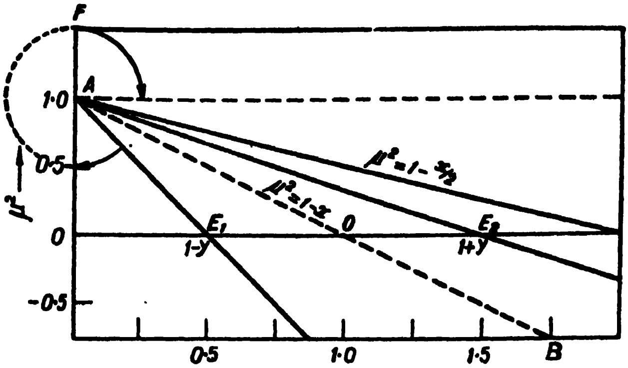


FIG. 8. Variation of μ^2 with x when the magnetic field is along the direction of propagation. (Case of longitudinal propagation.)

y>1

For the upper sign, the curve swings from AF' to AD as y increases from 1 to ∞. For the lower sign, it swings from AC to AD.

It will be noted that for $y<1$, it is possible for both the curves (corresponding to the upper and lower signs) to give values of $\mu^2 = 0$. For values of y greater than 1, the curve corresponding to the lower positive sign only can give values of $\mu^2 = 0$. For any value of y, $\mu^2 = 0$, when $x = 1-y$ and $x = 1+y$.

Magnetic field at right angles to the direction of propagation (transverse propagation)

In this case $\gamma_L = 0$ and $\gamma_T = \gamma$. We have from Eq. (22)

$$\mu^2 = 1-x \qquad \text{(upper positive sign)} \quad .. \quad .. \quad (25)$$

and

$$\mu^2 = 1-\frac{x(1-x)}{1-x-y^2} \qquad \text{(lower negative sign)}. .. \quad .. \quad (26)$$

Here again, we distinguish between the two cases $y<1$ and $y>1$.

y<1

The curve corresponding to the upper positive sign is the central straight line AB (Fig. 9)—which is the same as when there is no magnetic field. For the lower negative sign, the curve is as shown by the full lines. Proceeding towards increasing values of x the curve cuts the x-axis at $x=1-y$ and then branches off to $-\infty$ at $x=1-y^2$, and on crossing the line $x=1-y^2$ reappears from $+\infty$ on the right and cuts the x-axis again at $x=1+y$.

y>1

The upper positive sign again gives the straight line AB (Fig. 10). The lower negative sign gives the full line curve. It gives values of μ^2

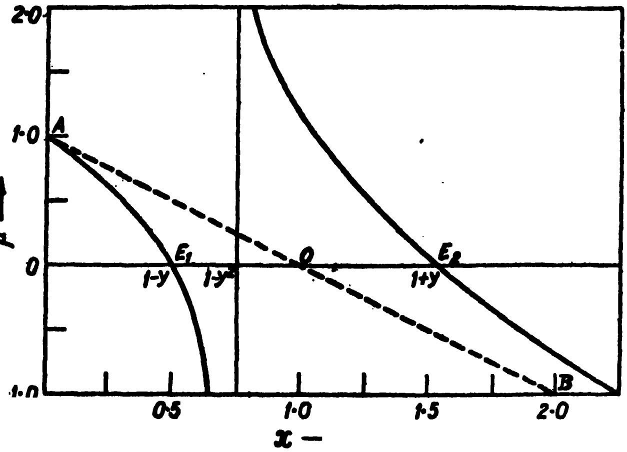


FIG. 9. Variation of μ^2 with x when the mgnetic field is at right angles to the direction of propagation. (Case of transverse propagation.) Case of $y<1$, i.e., $p>p_H$.

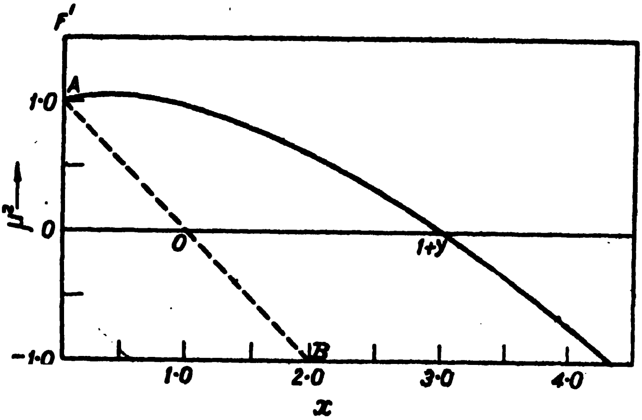


FIG. 10. Same as Fig. 9. Case of $y>1$, i.e., $p<p_H$.

greater than unity for values of x between 0 and 1 and cuts the x-axis at $x = 1+y$. It will be noted that there is no discontinuity in the curve as in the previous case.

Magnetic field inclined at any angle to the direction of propagation (general case)

Let the direction of propagation make an angle θ with the direction of the magnetic field. As before, we distinguish between the two cases $y<1$ and $y>1$.

$y<1$

A typical curve is shown in Fig. 11. The broken line portions are obtained by using the negative sign and the continuous ones by using the

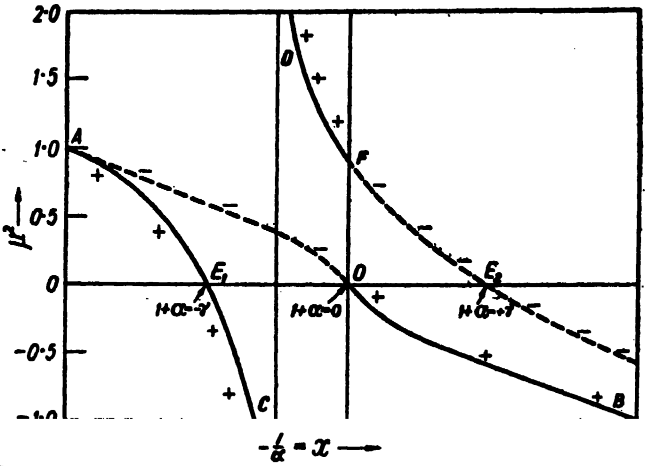


FIG. 11. Variation of μ^2 with x for an intermediate direction of propagation. The continuous portions of the curves are obtained by using the positive sign and the broken line portions by using the negative sign before the radical in Eq. (22). Case of $y<1$.

positive sign before the radical in Eq. (22). Taking the positive sign we note that the curve AE_1 branches off towards infinity at C and reappears at D. The asymptote DC passes through the point

$$x = -\frac{1}{\alpha} = \frac{\alpha^2-\gamma^2}{\alpha^2-\gamma^2\cos^2\theta}.$$

At F the curve gives a discontinuous jump as it were, and branches off towards the right from O along OB. The curve given by the negative sign has no infinity, but it also gives a discontinuous jump on crossing the vertical line through O $(1+\alpha = 0)$. It is thus seen that to maintain continuity of the curves on crossing this line one has to change sign before the radical. The reason of the curious jumps which occur whenever the curves cross the line $-1/\alpha = x = 1$ can be understood if it is remembered that to the left of the point O, $1+\alpha$ is negative, and to the right it is positive. These points have been fully discussed by Bhar [21]. The points at which the curve cuts the x-axis are given by

$$1+\alpha = -\gamma \quad \text{or,} \quad p_0^2 = p^2-pp_H \quad \quad (27)$$

$$1+\alpha = 0 \quad \text{or,} \quad p_0^2 = p^2 \quad \quad (28)$$

$$1+\alpha = \gamma \quad \text{or,} \quad p_0^2 = p^2+pp_H. \quad \quad (29)$$

The curve given by the positive sign is asymptotic to infinity at

$$x = \frac{\alpha^2-\gamma^2}{\alpha^2-\gamma^2\cos^2\theta}$$

or when

$$\frac{p_0^2}{p^2} = \frac{p^2-p_H^2}{p^2-p_L^2}. \quad \quad (30)$$

At the points E_1, O and E_2 represented by the Eqs. (27), (28) and (29), $\mu^2=0$ and $d\mu/dx=\infty$. For the point represented by Eq. (30), $d(\mu^2)/dx$ is also infinity. We shall presently see that these points give the conditions for which a wave incident normally on an ionospheric layer of gradually increasing electron density with height, may be totally reflected. The condition

$$x=\frac{\alpha^2-\gamma^2}{\alpha^2-\gamma^2\cos^2\theta}$$

is often called the *fourth condition* of reflection. We will refer to this condition again when discussing the Lorentz polarization term [Sec. 4(*f*)].

It is instructive to study how the shapes of the dispersion curves vary as the relative magnitudes of γ_T and γ_L gradually change, i.e., as the direction of propagation is gradually changed from, say, transverse to the magnetic field to along the magnetic field. This is shown in Fig. 12. For the transverse case the negative sign gives the straight line AO; its prolongation

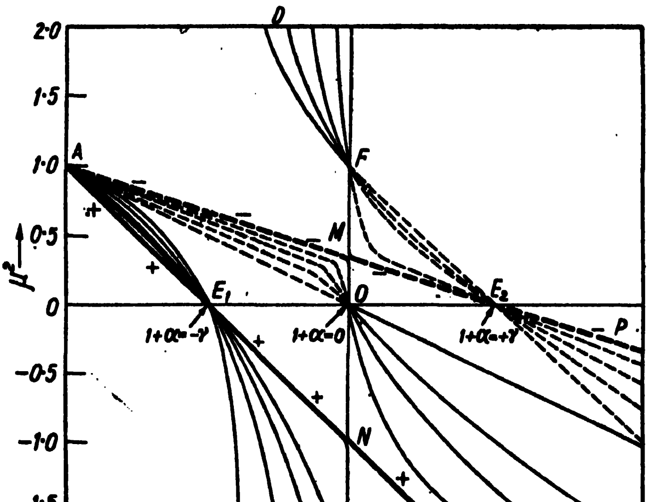


FIG. 12. Illustrating the transition of the dispersion curves from transverse to longitudinal mode of propagation and *vice versa*. The continuous portions are obtained by using the positive sign and the broken line portions by using the negative sign before the radical in Eq. (22). Case of $y<1$.

is, however, given by the positive sign as explained above. The positive sign gives the branch AE_1C which, after passing through infinity, reappears at D and gives DF; the prolongation of DF, namely FE_2, is, however, given by the negative sign. As the direction of propagation gradually alters, the curves change their shapes and when the direction of the magnetic field coincides with the direction of propagation they degenerate into two straight lines AME_2P (negative sign) and AE_1NQ (positive sign). At the same time the asymptote DC (Fig. 12) gradually shifts towards the line

OF and in the limit coincides with it. The dispersion curve in the form of two straight lines AME_2P and AE_1NQ was obtained by Lorentz for the special case of propagation along the magnetic field (longitudinal case) many years before the general formula was derived.

y>1

In Fig. 13 the transition from the transverse to the longitudinal mode of propagation is shown for the case $y>1$. The two thick lines AE_1Q and AME_2P in the figure relate to the case of propagation in the direction

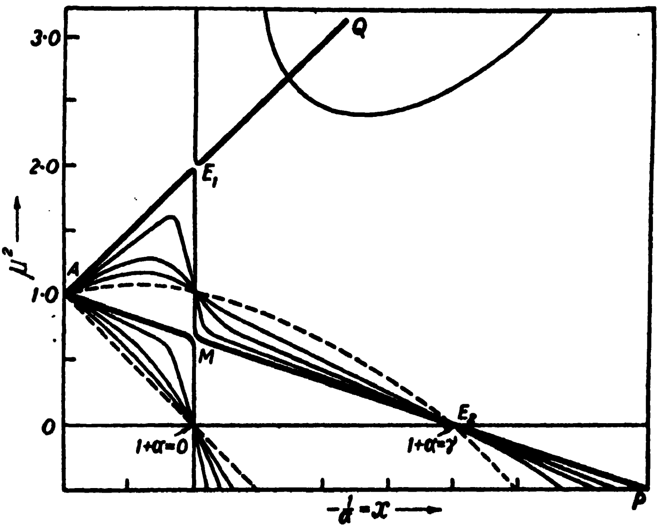


FIG. 13. Illustrating the transition of the dispersion curves from transverse to longitudinal mode of propagation and *vice versa*. Case of $y>1$.

of the magnetic field. These lines correspond to the two lines AE_1 and AE_2 of Fig. 12. It will be noticed that in this case, for any direction of propagation except strictly longitudinal, the dispersion curves cut the x-axis at two points only, namely, at

$$1+\alpha = 0, \quad \text{or,} \quad p_0^2 = p^2 \quad \ldots \quad \ldots \quad \ldots \quad (28)$$

and

$$1+\alpha = \gamma, \quad \text{or,} \quad p_0^2 = p^2+pp_H. \quad \ldots \quad \ldots \quad (29)$$

The dispersion curve for the wave which has a zero at $1+\alpha = 0$, has an infinity at

$$x = \frac{\alpha^2-\gamma^2}{\alpha^2-\gamma^2\cos^2\theta}.$$

For the longitudinal mode of propagation the curve cuts the x-axis at a single point, namely, $1+\alpha = \gamma$.

(ii) *Nomenclature of the split components: Ordinary and Extraordinary.*—A study of the dispersion curves discloses the remarkable fact that the values of x, for which μ^2 becomes equal to zero, are independent of the inclination of the magnetic field to the direction of propagation. Thus, if

we follow the curves in Fig. 12 for any inclination of the magnetic field to the direction of wave propagation (except when the two directions are actually coincident) it is seen that the refractive indices are always brought down to zero value at E_1, O and E_2. Now the electron number density corresponding to the point O is given by $1+\alpha = 0$ or $x=1$, i.e., the condition of reflection represented by this point is the same as that for the case of no magnetic field. The ray which undergoes reflection at this point is, therefore, called the ordinary ray since its condition of reflection is unaffected by the magnetic field.

The other ray, of which the refractive index varies along the curves AE_1 and FE_2 and attains zero value at the points $1+\alpha = \pm\gamma$, or $x = 1\pm y$, is called the extraordinary ray since its conditions of reflection (as represented by these points) are affected by the magnetic field.

This nomenclature has obviously been derived from the optical analogy. In double refraction in optics the ray which follows the ordinary laws of refraction is called the ordinary ray and the other is called the extraordinary ray. Unfortunately, however, the nomenclature in the ionospheric case is not quite as unambiguous as in the optical case [22]. For instance, for the case of propagation along the magnetic field for $y<1$ the two branches of the dispersion curve AE_1 and AE_2 in Fig. 8 give the value $\mu^2 = 0$ at $x=1\pm y$, i.e., both the conditions of reflection are affected by the magnetic field. Thus, logically, in such a case both the reflected rays should be called extraordinary. Again, in the case of oblique incidence it is found that the reflection conditions of *both* the magneto-ionic components are *always* affected by the magnetic field. One cannot therefore, according to the convention adopted above, call one ordinary and the other extraordinary. In this case the component which is less affected by the magnetic field, i.e., the one which in the limiting case of vertical propagation is not affected by the magnetic field, is called the ordinary and the other the extraordinary.

(iii) *Polarization.*—If we put $\beta=0$ and $x = -1/\alpha$, and $y=p_H/p$, as before, the expression for the polarization R becomes

$$R = \frac{h_z}{h_y} = \frac{-j}{\gamma_L}\left[\frac{1}{M^2-1} - \alpha\right] = -j\,[f \pm \sqrt{1+f^2}], \qquad .. \quad (31)$$

where

$$f = g \cdot \frac{1}{1-x}; \quad g = \frac{y}{2} \cdot \frac{\sin^2\theta}{\cos\theta},$$

θ being the angle between the direction of propagation and the positive direction of the magnetic field as shown in Fig. 14.

The sign before the expression for R has been made negative because the charged particles in the ionosphere which influence the propagation of waves are supposed to be electrons. R thus signifies the ratio between the components of the magnetic vector of the wave in and at right angles to the plane containing the wave normal and the direction of the magnetic field hereinafter called the magnetic plane. Thus, when R is zero, h_z is zero and the wave is plane polarized with its magnetic vector perpendicular

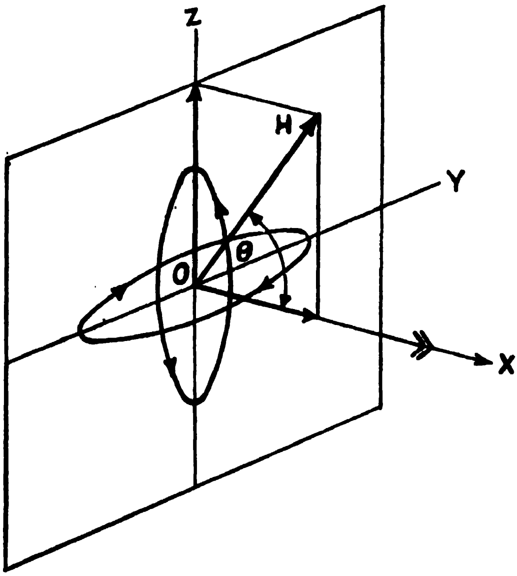


FIG. 14. Illustrating elliptic polarization of the magnetically split components.

to the magnetic plane. When R is infinite, h_y is zero and the wave is again plane polarized but with its magnetic vector in the magnetic plane. When R is numerically equal to unity, the magnitudes of h_z and h_y are equal and the wave is circularly polarized. The polarization is called right-handed when the rotation appears to be clockwise on looking towards the direction of propagation and left-handed when it appears anti-clockwise. Right-handed or left-handed circular polarization occurs according as h_z/h_y is $-j$ or $+j$.

Since the extraordinary ray is the one for which the condition of reflection is affected by the magnetic field $(1+\alpha = \pm\gamma)$ and the ordinary ray is that for which it is independent of the magnetic field $(1+\alpha = 0)$, it is evident from Fig. 12 that for values of x less than or equal to 1, the extraordinary ray is given by the upper positive sign before the radical in Eq. (31) and the ordinary ray by the lower negative sign; for values of x greater than 1, the signs have to be changed, i.e., the extraordinary ray is given by the lower negative sign while the ordinary by the upper positive sign.

From this it follows, according to Eq. (31), that if the direction of propagation makes an acute angle with the direction of the magnetic field, then the ordinary component is polarized in the left-handed and the extraordinary one in the right-handed sense. The case is reversed when the two directions include an obtuse angle. These results are important in the study of the ionosphere. In the northern hemisphere the direction of the field makes an acute angle with the direction of the downcoming waves. The polarizations of the split components, therefore, correspond

to the first case. For the southern hemisphere, the direction of the field makes an obtuse angle with the direction of the downcoming wave and the senses of polarization correspond to the second case. Without going into detailed analysis we will describe here the nature of polarization of the various types of waves that are propagated in the ionosphere. We will first deal with two special cases, namely, the longitudinal and the transverse, and then indicate in a general manner the nature of polarization for any direction of propagation.

Longitudinal case: In this case θ being equal to zero, $R = \mp j$; the two waves into which an incident wave is split up are both circularly polarized but in the opposite senses.

Referring to Fig. 12, we have seen how, when the transition occurs from the general case to the longitudinal case, the two lines AM and MP and also AN and NQ join up to form the two lines of the longitudinal case after Lorentz. At M and N there are no discontinuous changes of polarization, the polarization throughout AMP being left-handed and that throughout ANQ right-handed. (This case is similar to the case of an acute angle between the direction of propagation and that of the magnetic field.)

Transverse case: Here, putting $\theta = 90°$, we find R to be equal to $-j\infty$, or 0; these are interpreted as states of plane polarization. For transverse propagation, therefore, the two component waves into which an incident wave is split up are, in contrast to the longitudinal case, both plane polarized. If we consider propagation in the region to the left of O, (Fig. 12), i.e.,

$$N_e < \frac{mp^2}{4\pi e^2},$$

we may say that the ordinary ray (corresponding to curve AO) has its magnetic vector perpendicular to the magnetic plane while the extraordinary ray (corresponding to curve AE_1) has its magnetic vector parallel to the same.

General case: The state of polarization for any direction of propagation, other than longitudinal or transverse, is in general elliptic. The two ellipses corresponding to the two split components have the same ratio of the major to the minor axis. They are, however, oriented at right angles to each other and their senses of rotation are opposite (Fig. 14). Since, as has already been mentioned, the continuity of the dispersion curves for any direction of propagation except strictly longitudinal (Fig. 12) entails a change of sign at $1+\alpha = 0$ or, $x = 1$, the senses of rotation also abruptly change sign on crossing the line OF, i.e., the level of reflection of the ordinary ray. If, as in the transverse case, we consider the region to the left of OF, we may write for the states of polarization of the ordinary and the extraordinary as

$$R_o = \frac{h_z}{h_y} = -j[f-\sqrt{1+f^2}], \quad .. \quad .. \quad .. \quad (32)$$

$$R_x = \frac{h_z}{h_y} = -j[f+\sqrt{1+f^2}] \quad .. \quad .. \quad .. \quad (33)$$

The polarization is thus a function of f which again is a function of y, i.e., the ratio of the gyrofrequency to the wave frequency, of θ the inclination of the wave normal to the direction of the magnetic field and of x which is proportional to the electron number density N_e (for a given wave frequency).

In order to determine the polarization we have to take all these factors into account. This may be done by first plotting graphs connecting

$$g\left(=\frac{y}{2}\cdot\frac{\sin^2\theta}{\cos\theta}\right) \text{ and } \theta$$

for various values of y (Fig. 15) and then graphs connecting R and x for various values of g (Fig. 16).

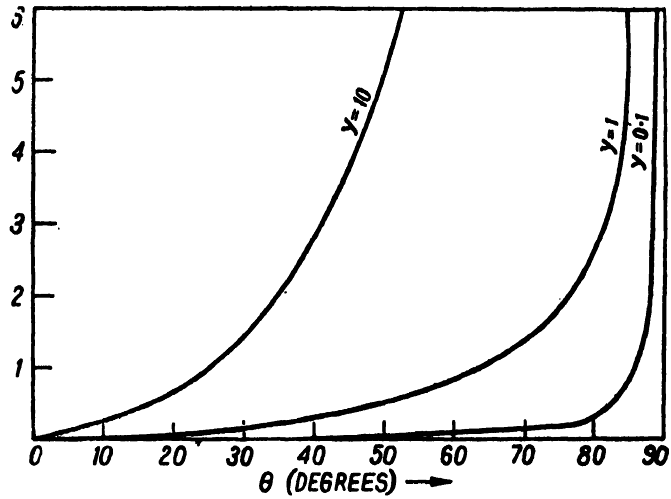


FIG. 15. Variation of $g\left(=\frac{y}{2}\cdot\frac{\sin^2\theta}{\cos\theta}\right)$ with θ for different values of y.

As an example, we deduce the state of polarization of a 150 m. wave travelling vertically downwards at Calcutta. (The value of λ_H at Calcutta is approximately 235·7 metres.) Here,

$$y = \frac{150}{235{\cdot}7} = 0{\cdot}63\,;$$

the magnetic dip being 31° 45′, $\theta = 58°\ 15'$. g is therefore equal to 0·437. The wave passing through an ionized layer of constant electronic density $N_e = 10^5$ per cm.^{3} will have for h_x/h_y the values $+0{\cdot}658j$ and $-1{\cdot}532j$ for the ordinary and the extraordinary rays respectively.

In actual ionospheric exploration by the echo method, as is described later, the polarization phenomenon is complicated by the fact that the wave travels through regions of continually changing electron number density and the state of polarization therefore changes continually, as is evident from Fig. 16 where R is plotted against x. If the exploring wave reaches the region of reflection of the ordinary wave where $N_e = mp^2/4\pi e^2$, i.e., $x = 1$,

the state of polarization changes sign, i.e., what was previously right-handed becomes left-handed and *vice versa*. We have already referred to this

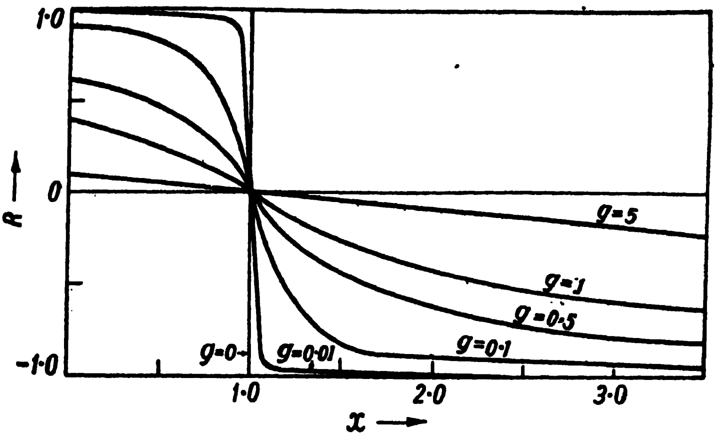


FIG. 16. Illustrating the changes in polarization as the wave of a given frequency enters into regions of increasing electron number density. R is the ratio of the major to the minor axis of the polarization ellipse. When the exploring wave reaches the region of reflection of the ordinary wave, i.e., $x=1$, the state of polarization changes sign.

sudden change of polarization at the point $x=1$, $R=0$, where the wave is plane polarized.

(*d*) Complete Appleton-Hartree formula including collisional term

We have till now discussed the propagation of the waves in the ionosphere without taking into account the effect of collisions of electrons and ions with neutral gas molecules and atoms. If collision is taken into account, then, under certain circumstances, its presence, in addition to causing absorption of the wave, profoundly affects the dispersion and the state of polarization [22*a*] of the wave.

The evaluation of μ, k and R from the complete complex equations (20) and (21) for the case of propagation in any direction with respect to the magnetic field is extremely complicated. The problem has been tackled in great detail by Mary Taylor [23], by Goubau [24] and by Westfold [24*a*], who have published a large number of graphs connecting N_e with μ, k and R. The derivation of these curves is laborious. Bailey [25] has also discussed these equations and developed an elegant graphical method of delineating them. By utilizing this method Martyn [26] has drawn dispersion, absorption and polarization curves for the value of the magnetic field in the southern hemisphere in Australia.

(i) *Dispersion.*—The nature of the dispersion is best understood from a study of the dispersion curves for certain typical cases. The full line curves in Figs. 17, 18 and 19, which are for the condition of Calcutta, have been drawn by Ghosh [27] after the method indicated by Bailey and Martyn.

As in the case of simple dispersion curves already discussed, the curves obtained by taking collision into account may also be classified into two

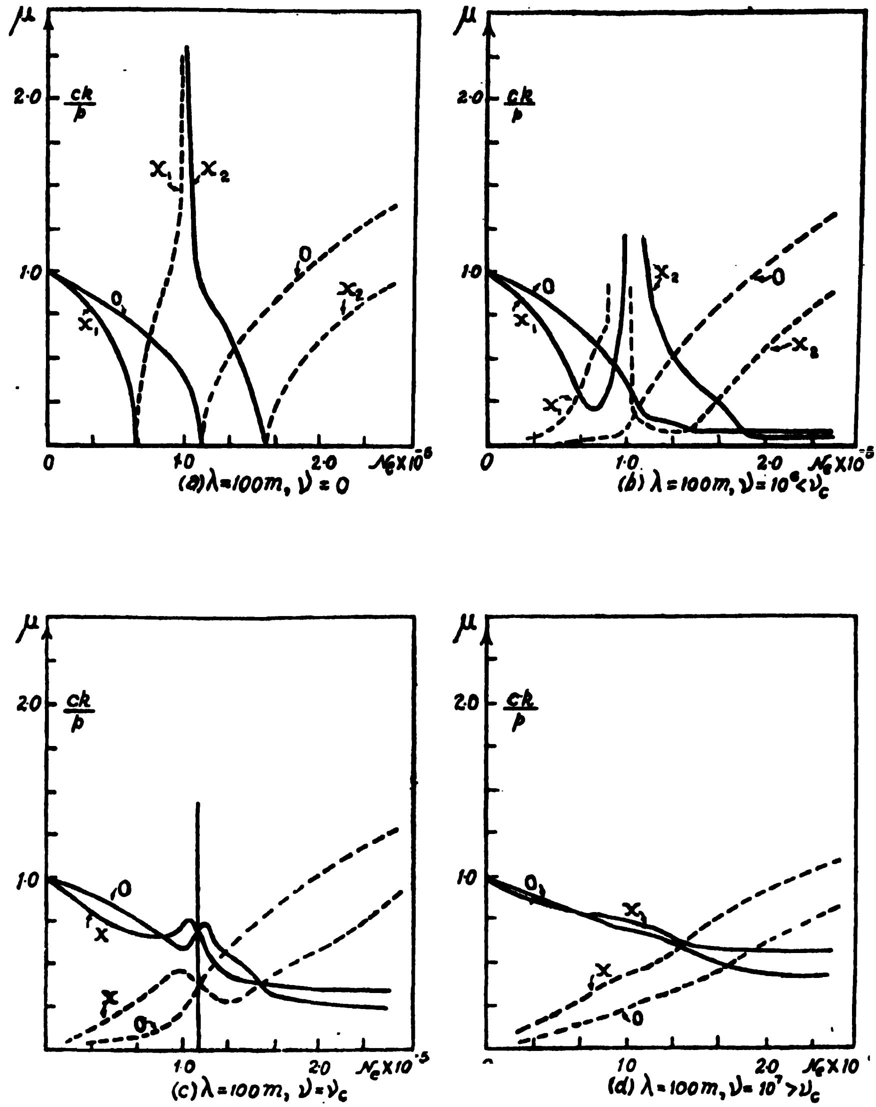


FIG. 17. Dispersion and absorption curves taking collisional frequency into account. The continuous curves represent the variation of μ and the broken line curves that of ck/p with N_e. Curves marked O correspond to the ordinary component and those marked X to the extraordinary component. The vertical line in (c) corresponds to $1+\alpha = 0$. The curves are for Calcutta ($\lambda_H = 235{\cdot}7$ metres). Case of $\lambda < \lambda_H$.

distinct types, those for which $p > p_H$ (Fig. 17) and those for which $p < p_H$ (Fig. 19). Fig. 18 is for the case of $p = p_H$. The curves are drawn for a set of increasing values of ν starting from $\nu = 0$, i.e., the simple case of no collision. Each of the types may be classified into two distinct cases, one

in which the nature of the curves is roughly similar to that of the simple curves for longitudinal propagation with no collision and the other, in which

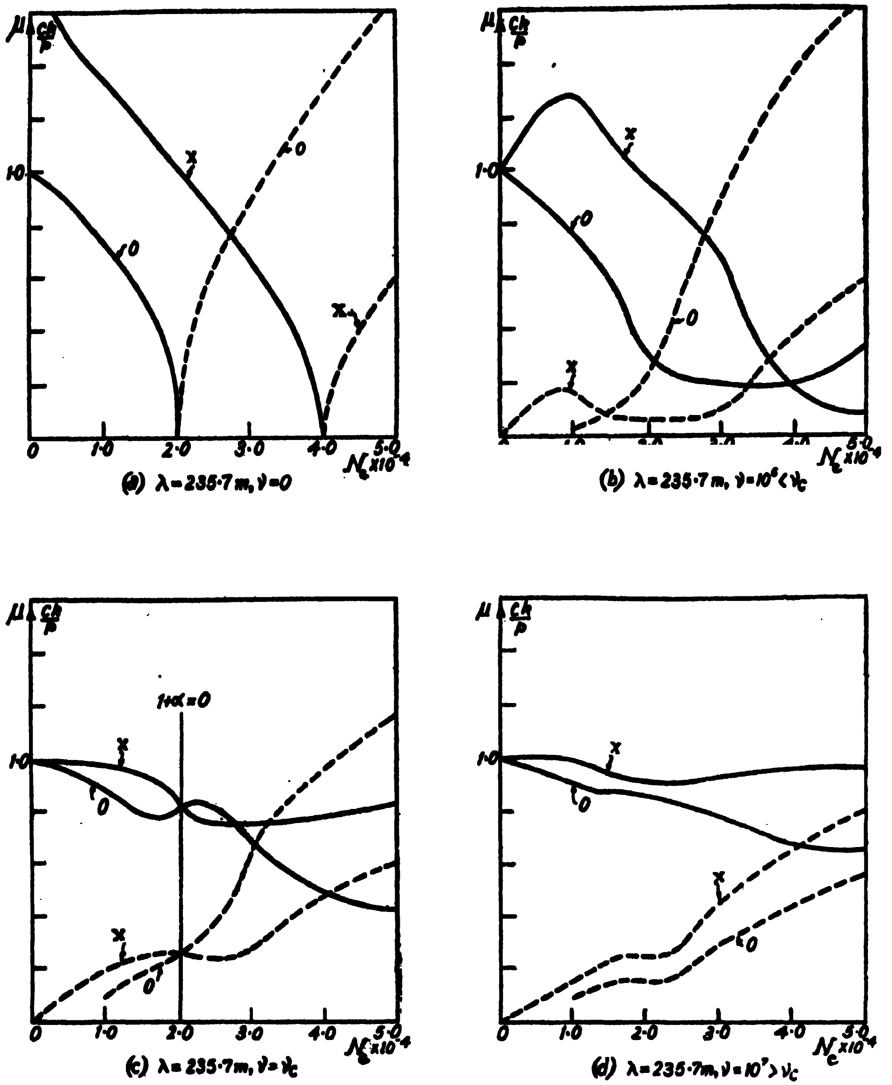


FIG. 18. Same as Fig. 17. Case of $\lambda = \lambda_H$.

the nature of the curves is roughly similar to the simple curves for transverse propagation with no collision. The former is called quasi-longitudinal and the latter quasi-transverse propagation. Curves (*d*) in Figs. 17, 18 and 19 belong to the quasi-longitudinal case while curves (*b*) in the same figures belong to the quasi-transverse case. Curves (*c*) represent the transition from one case to the other; the condition under which this takes place is discussed below.

It will be noticed that the curves in (*b*) can be derived roughly from those in (*a*) (which are simple dispersion curves with no collision) if it is

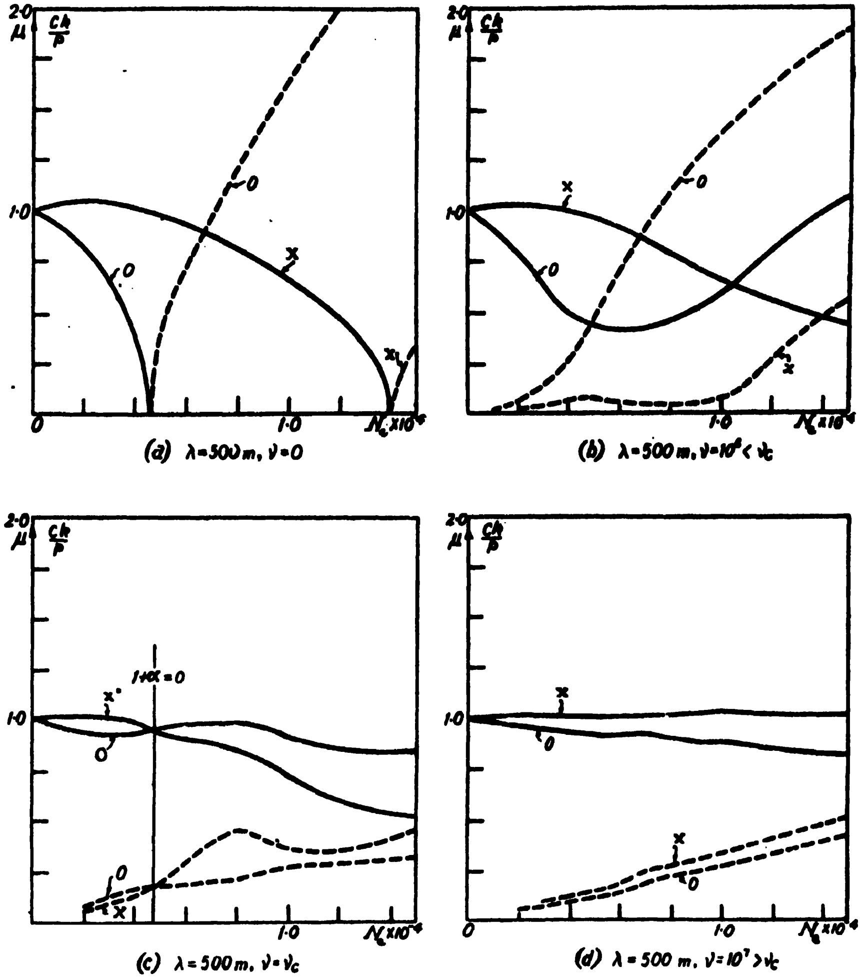


FIG. 19. Same as Fig. 17. Case of $\lambda > \lambda_H$.

imagined that the infinities have been removed and the two branches of the extraordinary have joined up. The curves instead of running to negative values of μ branch off towards the right and proceed more or less parallel to the N_e-axis. (For very high values of N_e, the curves again tend to rise.) In (*d*), however, the shapes of the curves have altered completely, and except for the two kinks, the curves tend to the two straight lines for the case of longitudinal propagation.

Now let us recall that for the case of no collision, a change-over from the transverse case to the longitudinal case occurs when the direction of wave propagation, instead of being inclined to the direction of the magnetic field, is actually coincident with it. The characteristic feature of the

change-over is that the ordinary branch joins up with the extraordinary and *vice versa*, across the line $x=1$ (i.e., at the level of ionization where $\mu = 0$ for the ordinary ray) and form two continuous dispersion curves.

Similar change-over occurs even when collision is present, across the line $1+\alpha=0$ $(x=1)$; but, in this case the direction of propagation need not coincide with the direction of the magnetic field. Instead, the condition for change-over, across the line $x = 1$ for any direction of propagation, is given by $\nu > \nu_c$ where ν_c is equal to $p_T^2/2p_L$. An inspection of the curves in (c) will show how the branch to the left of the line $x = 1$, which was originally ordinary, has joined up with the branch on the right, which had been extraordinary and *vice versa*. If $\nu < \nu_c$, the dispersion curves across this line behave more or less like the dispersion curves in the transverse case (Fig. 12); this may therefore be called *quasi-transverse* type of propagation. If $\nu > \nu_c$, the dispersion curves across this line resemble those of the longitudinal case; this may therefore be called *quasi-longitudinal* type of propagation.

If we consider the dispersion curves as a whole, instead of considering the change-over that takes place across the line $x=1$, we may conveniently derive from Eq. (20) the conditions for quasi-transverse and quasi-longitudinal types of propagation. Considering the terms within the radical in Eq. (20) it is seen that if

$$\left|\frac{\gamma_T^4}{4(1+\alpha+j\beta)^2}\right| > \left|\gamma_L^2\right|,$$

M, the complex refractive index, depends on γ_T alone. Under this condition we call the propagation *quasi-transverse*. Again, when

$$\left|\gamma_L^2\right| > \left|\frac{\gamma_T^4}{4(1+\alpha+j\beta)^2}\right|,$$

M depends on γ_L alone. Under this condition the propagation is called *quasi-longitudinal*.

It is easy to see that since along the line $x=1$, $1+\alpha = 0$, the condition of quasi-transverse or quasi-longitudinal propagation is determined by

$$\left|\gamma_T^4/4\beta^2\right| > \text{ or } < \left|\gamma_L^2\right|, \text{ i.e., } \nu < \text{ or } > \nu_c \, (= p_T^2/2p_L).$$

(ii) *Absorption*.—The expression for the complex refractive index, $M = \mu - jck/p$, shows that the refractive index and absorption coefficient are closely related to one another. In Figs. 17–19 the broken line curves representing absorption in a length of path $\lambda/2\pi$, i.e., ck/p are drawn for cases corresponding to the dispersion curves depicted therein. It will be noticed that the absorption tends to a maximum as the refractive index tends to a minimum value.

We may first discuss the absorption for the particular and comparatively simple case of longitudinal propagation, i.e., when $\gamma_T = 0$. From

Eq. (20) we find that when $\gamma_T = 0$, and when squares and higher powers of ck/p can be neglected

$$\mu^2 = 1 - \frac{p_0^2(p \mp p_L)}{p[(p \mp p_L)^2 + \nu^2]} \quad \ldots \quad (34)$$

$$k = \frac{\nu}{2c} \frac{p^2 p_0^2}{\sqrt{[(p^2 \mp pp_L)^3 (p^2 - p_0^2 \mp pp_L)]}} \cdot \quad \ldots \quad (35)$$

If p_0 is small compared to p, the expression for k is further simplified to

$$k = \frac{\nu}{2c} \cdot \frac{p_0^2}{(p \mp p_L)^2}$$

and the corresponding absorption curves are given in Fig. 20. It is seen that k tends to infinity as μ^2 approaches zero value.

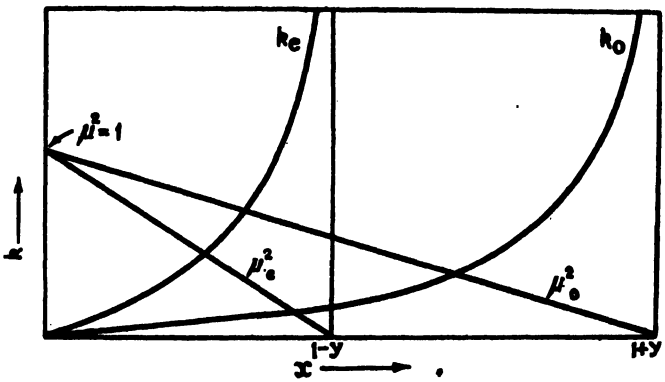


FIG. 20. Variation of the absorption coefficient k for the case of longitudinal propagation.

We will now consider the quasi-longitudinal and the quasi-transverse cases of propagation. These are the types of propagation more often met with in practice; the collisional frequency ν is more likely to be greater or less than ν_c $(= p_T^2/2p_L)$ than to be equal to it. The expressions for these special cases are also much simpler [28] than the general expression in Eq. (20).

Eq. (20) may be put in the form

$$M^2 = \left(\mu - \frac{jck}{p}\right)^2 = 1 + \frac{1}{X + jY} \cdot$$

Separating the real and the imaginary parts we get

$$\mu^2 - \frac{c^2k^2}{p^2} = 1 + \frac{X}{X^2 + Y^2} \quad \ldots \quad (36)$$

and

$$\frac{2\mu ck}{p} = \frac{Y}{X^2 + Y^2} \quad \ldots \quad (37)$$

where

$$X+jY=\alpha+j\beta-\frac{\gamma_T^2}{2(1+\alpha+j\beta)}\pm\sqrt{\left[\frac{\gamma_T^4}{4(1+\alpha+j\beta)^2}+\gamma_L^2\right]}. \quad .. \quad (38)$$

As already mentioned the propagation is quasi-transverse or quasi-longitudinal according as

$$\left|\frac{\gamma_T^4}{4(1+\alpha+j\beta)^2}\right| > \text{ or } < \left|\gamma_L^2\right|.$$

Introducing these conditions in the expression for $X+jY$ we find that for the *quasi-transverse* case:

$$\text{(i)}\quad \left.\begin{aligned} X&=\alpha \\ Y&=\beta \end{aligned}\right\}\text{ upper positive sign (ordinary)} \quad .. \quad .. \quad .. \quad (39)$$

$$\text{(ii)}\quad \left.\begin{aligned} X&=\alpha-\frac{\gamma_T^2(1+\alpha)}{(1+\alpha)^2+\beta^2} \\ Y&=\beta\left[1+\frac{\gamma_T^2}{(1+\alpha)^2+\beta^2}\right] \end{aligned}\right\}\begin{array}{c}\text{lower negative sign} \\ \text{(extraordinary)}\end{array} \quad .. \quad .. \quad (40)$$

and for the *quasi-longitudinal* case:

$$\text{(i)}\quad \left.\begin{aligned} X&=\alpha+|\gamma_L| \\ Y&=\beta \end{aligned}\right\}\begin{array}{c}\text{upper positive sign} \\ \text{(extraordinary)}\end{array} \quad .. \quad .. \quad .. \quad (41)$$

$$\text{(ii)}\quad \left.\begin{aligned} X&=\alpha-|\gamma_L| \\ Y&=\beta \end{aligned}\right\}\begin{array}{c}\text{lower negative sign} \\ \text{(ordinary)}\end{array} \quad .. \quad .. \quad .. \quad (42)$$

The values of absorption coefficient for the two cases of propagation, quasi-longitudinal and quasi-transverse, may now be deduced for both ordinary and extraordinary waves by substituting the appropriate values of X and Y from above in Eqs. (36) and (37).

In estimating the absorption suffered by a wave in its passage through an ionized medium, it is convenient to consider the absorptions separately in the regions in which the wave suffers little deviation, i.e., μ is nearly equal to 1, and the region where it suffers appreciable deviation, i.e., near the region of reflection where $\mu<1$ [29]. The first, we call the *non-deviating region* and the second the *deviating region*.

In the non-deviating region μ is nearly equal to unity. Hence from Eq. (37) the absorption per $\lambda/2\pi$ becomes

$$\frac{ck}{p}=\frac{Y}{2\mu(X^2+Y^2)}=\frac{1}{2}\frac{Y}{X^2+Y^2}. \quad .. \quad (37.1)$$

In the deviating region μ varies from point to point. If we neglect $(ck/p)^2$, then from Eqs. (36) and (37) we get

$$k=-\frac{1}{2}\cdot\frac{p}{c}\cdot\frac{Y}{X}\left(\frac{1}{\mu}-\mu\right). \quad .. \quad .. \quad .. \quad (43)$$

In order to find out the total absorption in any region we have to obtain the value of the integral $\int k\,dh$, where dh is an element of height.

For this, we first determine whether the propagation is of the quasi-transverse or of the quasi-longitudinal type. The appropriate values of X and Y are then substituted in the expression for k and the integration carried out.

(iii) *Polarization.*—The effect of damping on the state of polarization of the downcoming waves is to rotate the principal axes of the polarization ellipse [22a]. This can be easily seen in Figs. 21–23 where the full line curves

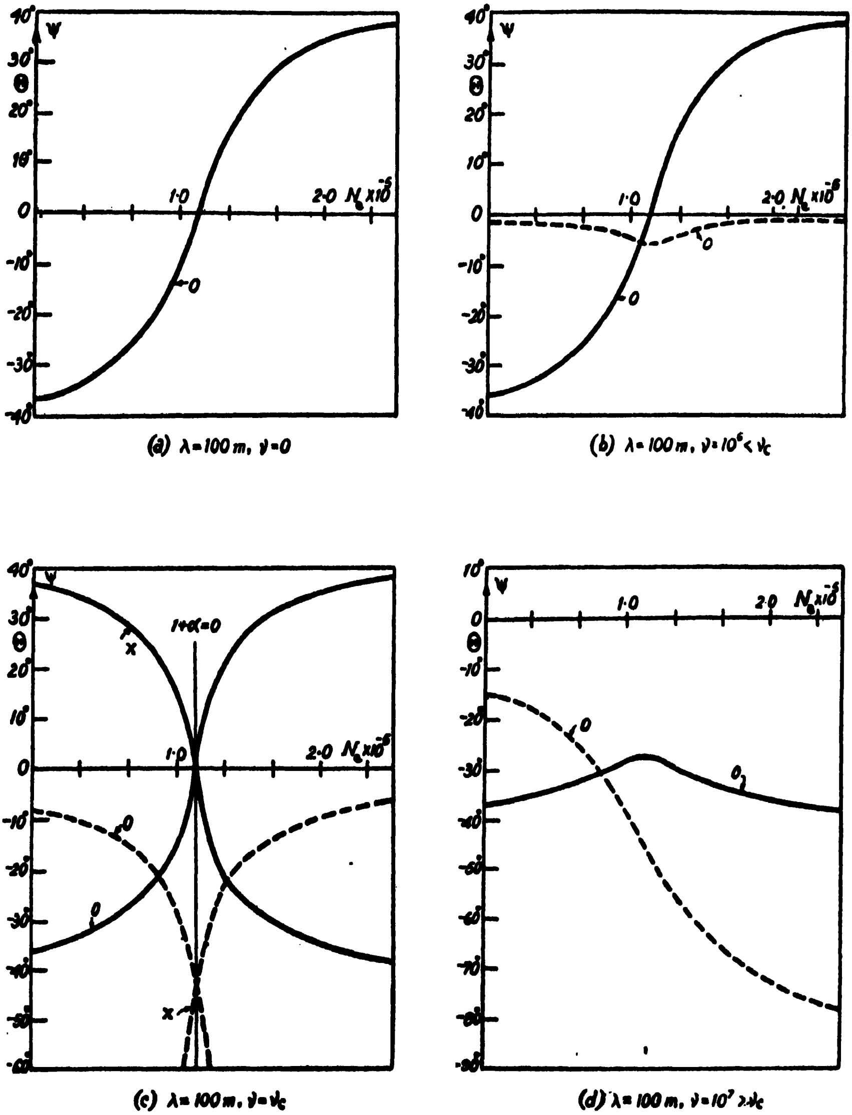


FIG. 21. Variation of the state of polarization Θ and of the tilt of the polarization ellipse Ψ for different values of the collisional frequency ν as the wave enters into regions of increasing electron number density. The curves are for Calcutta ($\lambda_H = 235{\cdot}7$ metres). Case of $\lambda < \lambda_H$.

represent Θ, the state of polarization, as defined by the relation $\tan\Theta = -j\,h_z/h_y$. The broken line curves represent Ψ, the tilt of the polarization ellipse, i.e., the inclination of the polarization ellipse to the Z-axis, for

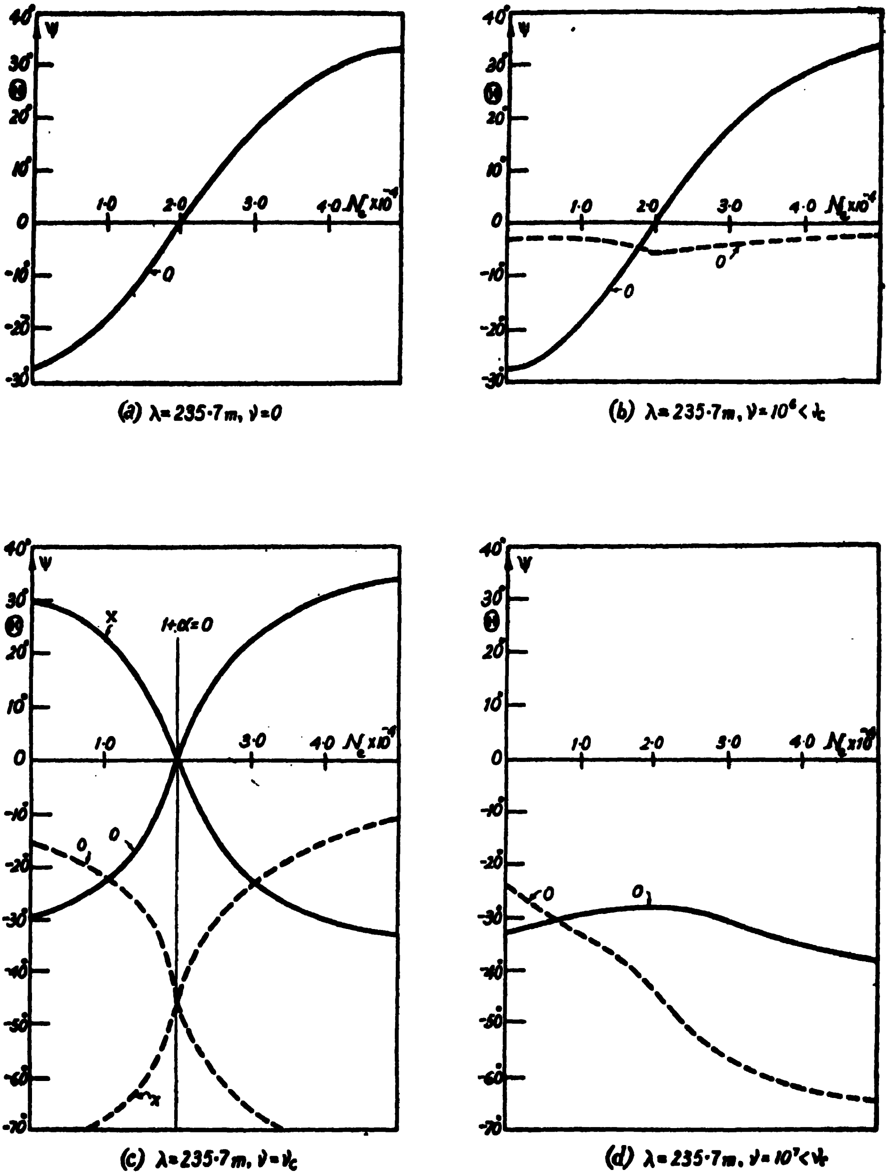


FIG. 22. Same as Fig. 21. Case of $\lambda = \lambda_H$.

the ordinary wave [27]. The curves have been drawn by Ghosh after the method of Bailey. The positive values of Θ correspond to right-handed polarization in the northern hemisphere. $\Theta = 0$ corresponds to linear polarization and $\Theta = \pm 45°$ corresponds to right- and left-handed circular

polarization. The value of Ψ is always negative and small for the case of quasi-transverse mode of propagation except at the point $1+\alpha=0$, where it attains maximum value.

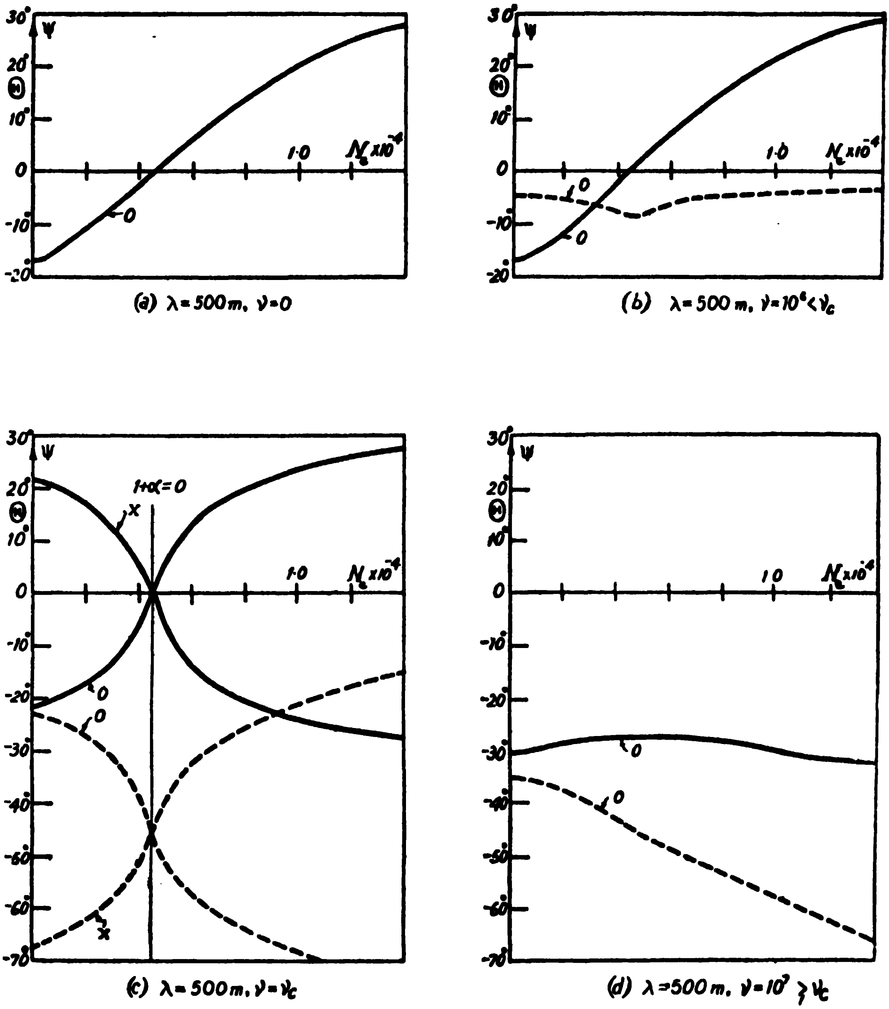


Fig. 23. Same as Fig. 21. Case of $\lambda > \lambda_H$.

It will be noted in Figs. 21–23 that for $\nu < \nu_c$, i.e., for quasi-transverse mode of propagation (curves in (*b*)) there is a change of the sense of polarization at the point $1+\alpha=0$; when $\nu > \nu_c$, i.e., for quasi-longitudinal propagation (curves in (*d*)) there is no such change and the nature of polarization continues to be the same for all values of N_e. The change-over from quasi-transverse to quasi-longitudinal type is best illustrated by the curves in (*c*).

Limiting polarization.—In all measurements of polarization of downcoming waves the nature of polarization studied is that which the waves

possess immediately on their emergence from the ionosphere. It is, therefore, important to consider what the polarization of the magnetically split components would be when they leave the ionosphere [17, 30].

In the case of vertical sounding of the ionosphere by radio waves, the magnetic field of the earth is, in general, inclined to the direction of wave propagation. A plane polarized wave, immediately on entry into the ionosphere, is therefore split up into two elliptically polarized components with opposite senses of rotation. These components, if they could retrace their original paths, would on emergence have that polarization which they possessed on entry at the boundary of the ionosphere, i.e., they would be elliptically polarized. This is exactly what the curves (Figs. 21–23) drawn after the method of Bailey show for the case of limiting polarization. It will be seen that as $N_e \to 0$, the values of Θ and Ψ (the polarization and the tilt of the polarization ellipse respectively) tend to limiting values which depend upon the angle between the direction of propagation and the direction of the magnetic field [26].

(e) Oblique incidence: Group path and phase path

The Appleton-Hartree magneto-ionic theory is particularly suitable for propagation of radio waves incident vertically on a horizontally stratified ionosphere. This is evident from the dispersion curves (Figs. 17–19) depicting variation of μ with N_e in which, amongst other parameters, the angle θ between the direction of propagation and the magnetic field is constant. For the case of vertical propagation this condition is practically satisfied. For oblique incidence, however, θ is no longer constant, but varies with μ, which depends in a complicated manner upon an unknown angle of refraction. This complication may be avoided by a generalization of the magneto-ionic theory, after Booker, in which the propagation of a magneto-ionic component is described in terms of the vertical component q of the phase-propagation vector (*vide infra*) instead of in terms of μ [31]. This generalized treatment incidentally brings out clearly certain special features of wave propagation in a stratified doubly-refracting medium.

The starting point in the analysis is the wave function

$$\exp\left[j\frac{2\pi}{\lambda}\{ct-\mu_\psi(y\sin\psi+z\cos\psi)\}\right] \quad \ldots \quad \ldots \quad (a)$$

which we have already referred to.

Here,

λ—wavelength in vacuo,

ψ—angle made by the direction of the ray with the vertical at the point (y, z),

and μ_ψ—value of the refractive index of the medium at this point (see Fig. 3).

Further, as before, the propagation is assumed to be in the vertical plane YZ, the Z-axis being vertical and the ionosphere is supposed to be

horizontally stratified with the electron density increasing with height. μ_ψ is a function of θ, N_e, H, ν and p, where the symbols have the same significance as in Sec. 2(*b*) and (*c*).

Instead of expression (*a*), Appleton [15], in deriving his magneto-ionic formula, has used the simpler wave function

$$\exp\left[j\frac{2\pi}{\lambda}(ct-qz)\right] \quad \cdots \quad \cdots \quad \cdots \quad (b)$$

in which the implicit assumption is made that the direction of wave propagation makes a constant angle with that of the magnetic field. This will be the case if the wave is propagated in a medium of constant electron density or in a stratified medium along the direction of the ionization gradient. In the form of the wave function (*a*) adopted by Booker, the element of path in the direction of the ray, which need not lie along the direction of the ionization gradient, is resolved into two components, one along the direction of the gradient and the other at right angles to the same, i.e., along the direction in which N_e is constant. Since, by Snell's Law, $\mu_\psi \sin\psi$ is a constant quantity and is equal to $\sin i$, where i is the angle which the direction of wave propagation makes with the vertical before the wave enters the ionized medium, the wave function takes the form

$$\exp\left[j\frac{2\pi}{\lambda}\{ct-y\sin i-qz\}\right], \quad \cdots \quad (c)$$

where q ($=\mu_\psi \cos\psi$) is a function of N_e, H, ψ and p. When the wave is incident on the ionosphere vertically, i.e., along the direction of ionization gradient, $i=0$, and expression (*c*) reduces to the simple form (*b*); q is then identical with the refractive index μ.

The method by which the expression for q is deduced may be recalled here. Wave function (*b*) is substituted in the Maxwellian equations for a medium containing free electrons in the presence of an external magnetic field. When this is done a quadratic equation in q^2 is obtained as

$$Aq^4+Cq^2+E=0, \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad (44)$$

where $A = (\alpha+j\beta)^2(1+\alpha+j\beta)-(\alpha+j\beta)\gamma_T^2-(1+\alpha+j\beta)\gamma_L^2$,

$C = 2(\alpha+j\beta)\gamma^2-2(\alpha+j\beta)(1+\alpha+j\beta)^2+\gamma_T^2+2\gamma_L^2$,

and $E = (1+\alpha+j\beta)^3-(1+\alpha+j\beta)\gamma^2$.

If, instead of the simple wave function (*b*), the wave function (*a*) is substituted in the Maxwellian equations, we get, instead of a quadratic in q^2, a quartic equation in q of the form

$$Aq^4+Bq^3+Cq^2+Dq+E=0. \quad \cdots \quad \cdots \quad \cdots \quad (45)$$

The coefficients A, B, C, D and E are in general complicated functions of N_e, p, θ, H and ν, but, for the particular case of vertical propagation in the ionosphere, coefficients B and D vanish and the equation becomes identical with (44).

To simplify the discussion, we will neglect the effect of collision, i.e., we assume ν, and, as a consequence, the damping to be equal to zero.

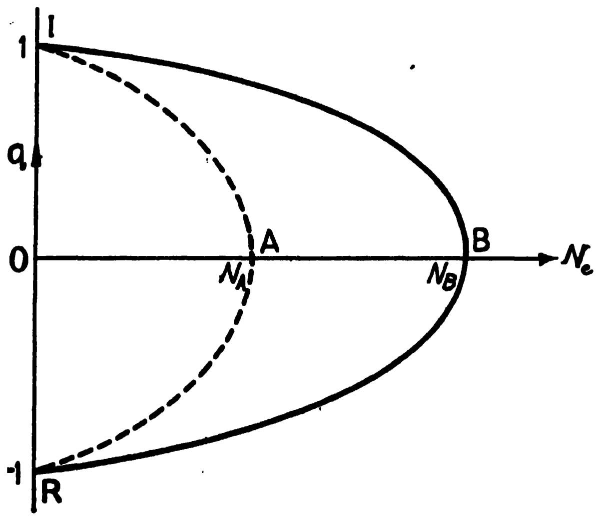


FIG. 24. Variation of q with N_e for the upgoing and downcoming waves. Case of vertical propagation. The full line curve is for the ordinary component and the broken line curve for the extraordinary component. Note that the curves cut the N_e-axis and turn towards the q-axis at the same points A and B [cf. Fig. 25]. (After Booker.)

The difference in the significances of the two equations (44) and (45) becomes apparent if we plot q (the four roots of the equations) as a function of N_e (Figs. 24 and 25). As mentioned before, q is not to be confused with the refractive index; the two quantities are identical only when $i=0$. Fig. 24 represents the values of q, as given by Eq. (44), plotted against N_e. The plot gives four curves IA, IB and AR, BR corresponding to the two pairs of equal and opposite roots, symmetrical about the N_e-axis. The curves IA and IB represent the upgoing magneto-ionic components, extraordinary and ordinary, and AR and BR the corresponding downcoming components. At the point A, corresponding to the electron density N_A, the two extraordinary roots become equal to one another. This means that reflection of the extraordinary wave will take place at the level of the ionosphere where the electron density is N_A. It will be seen that at this level not only is q equal to zero but also $dq/dN_e = \infty$. Similarly, the ordinary wave is reflected at the level of the ionosphere where the electron density is equal to N_B. At this level again $q=0$ and $dq/dN_e = \infty$.

Let us now interpret Fig. 25 which is a plot for the case of oblique incidence, i.e., of the roots of Eq. (45). Below the ionosphere where $N_e=0$, q is equal to $\cos i$ for the upgoing wave and to $-\cos i$ for the emergent downcoming wave [Sec. 2(*a*)]. The points of incidence and emergence of the wave are I and R. On entry into the ionosphere, the wave is split

into the two magneto-ionic components—ordinary and extraordinary. Curve *IDAR* represents the propagation of the extraordinary and curve

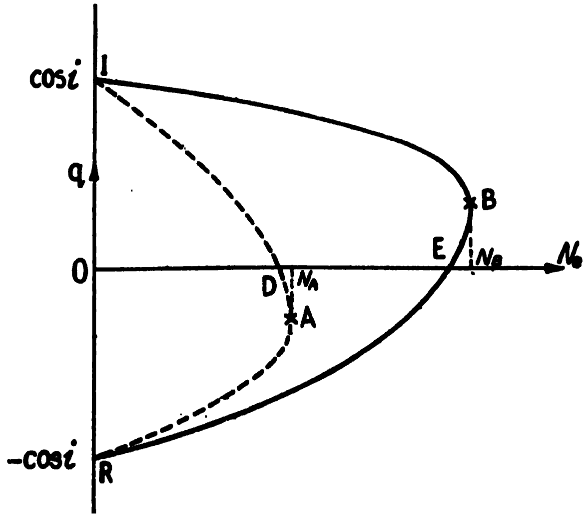


FIG. 25. Same as Fig. 24. Case of oblique propagation. Note that unlike Fig. 24, the points where the curves turn towards the q-axis are different from those where they cut the N_e-axis. (After Booker.)

IBER that of the ordinary. As the wave proceeds upwards into regions of increasing electron density, q decreases and attains zero value at D and E. The most striking point to be noted here is that though at D and E, q becomes equal to zero, yet these are not the points of reflection. The point of reflection for the extraordinary wave is at A where the curve *IDAR* begins to turn back towards smaller electron densities. Similarly, B is the point of reflection for the ordinary ray. It is at these points, A and B, that $dq/dN_e = \infty$.

We may use the $q-N_e$ curves of Fig. 25 for tracing the direction of propagation of the phase-ray. Since $\mu_\psi \sin \psi = \sin i$ = constant, a reference to Fig. 4 shows that a decrease in q means that ψ increases, i.e., the direction of the ray bends round towards the horizontal. We thus obtain the portion *id* in Fig. 26(*a*) corresponding to *ID* of Fig. 25; the direction of the phase-ray becomes horizontal at *d*, corresponding to the point D in Fig. 25. From D to A, the electron density is still increasing but q has now a negative value. Hence, corresponding to this portion of the curve the direction of phase-propagation must be downward as shown by *da* in Fig. 26(*a*). From A (Fig. 25), the curve *IDAR* turns towards regions of lower electron densities and the corresponding phase-propagation is represented by *ar* in Fig. 26(*a*). Thus the phase-propagation of the extraordinary wave in the ionosphere is completely shown by the curve *idar*. Fig. 26(*b*) shows the path of the phase-ray corresponding to the ordinary wave. Schekulin [32] was the first to draw attention to this peculiar nature

of phase-propagation. He, however, did not offer a detailed explanation of this. It is now seen how, from the $q-N_e$ curves, the paths of the phase-rays can be deduced.

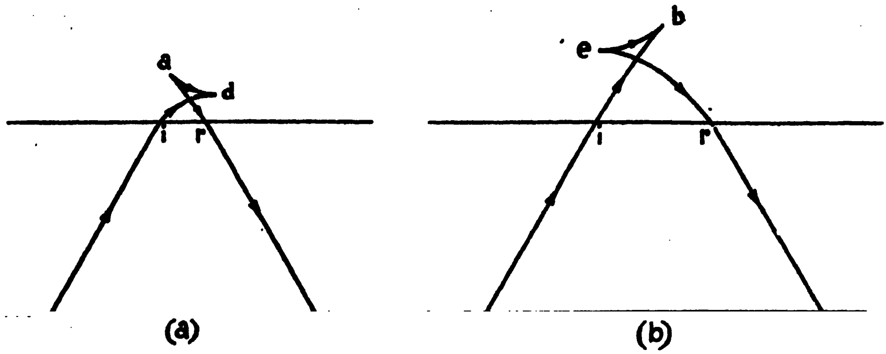


FIG. 26. Illustrating the paths of the phase-rays corresponding to the two magneto-ionic components. (*a*) Extraordinary ray, and (*b*) ordinary ray. (After Booker.)

It is to be noted that though at *d* and *e* (Fig. 26), $q=0$, and the direction of the phase-propagation becomes horizontal, yet these are *not* the points where the wave is reflected. The general condition of reflection, $dq/dN_e = \infty$, occurs at *a* and *b*. It can be shown after Booker that if, as in ionospheric exploration, a 'pulse' or wave-group travels upwards then these are precisely the points where the direction of group-propagation of the two component waves will become horizontal. In other words, the wave-groups as a whole would suffer total reflection at these two points. It is obvious that for the case of vertical incidence the group-velocity would become zero at the point of reflection.

The difference in the behaviours of the wave-group and the individual waves is due to the well-known fact that in a doubly-refracting medium, the direction of phase-propagation does not in general coincide with that of group-propagation [32*a*, 32*b*]. It is interesting and instructive to compare the path of a wave-packet with that of the individual waves of the packet, i.e., with the phase-path. Fig. 27 shows the group- and phase-rays of the two magneto-ionic components—(*a*) for the extraordinary, and (*b*) for the ordinary wave. It is seen that though the phase-rays have curious cusps as described above, the group-paths have no such cusps and are of the simple conventional type. The direction of propagation of the group as a whole follows the direction of the curve on the left-hand side. The small arrows indicate the direction of propagation of individual waves at different points of the path. These, it will be noticed, are parallel to the directions of the phase-rays at the corresponding points on the right-hand-side curves. In the lower portion of the curves, below the doubly-refracting and stratified ionosphere, the group-paths and the phase-paths are coincident with one another.

An important point which has got a bearing on the exploration of the ionosphere by vertically propagated 'pulses' is that in this particular

case, the group-paths and the phase-paths coincide and that, as a consequence, the point where the direction of the phase-rays becomes horizontal, is also the point where the direction of group-propagation becomes horizontal.

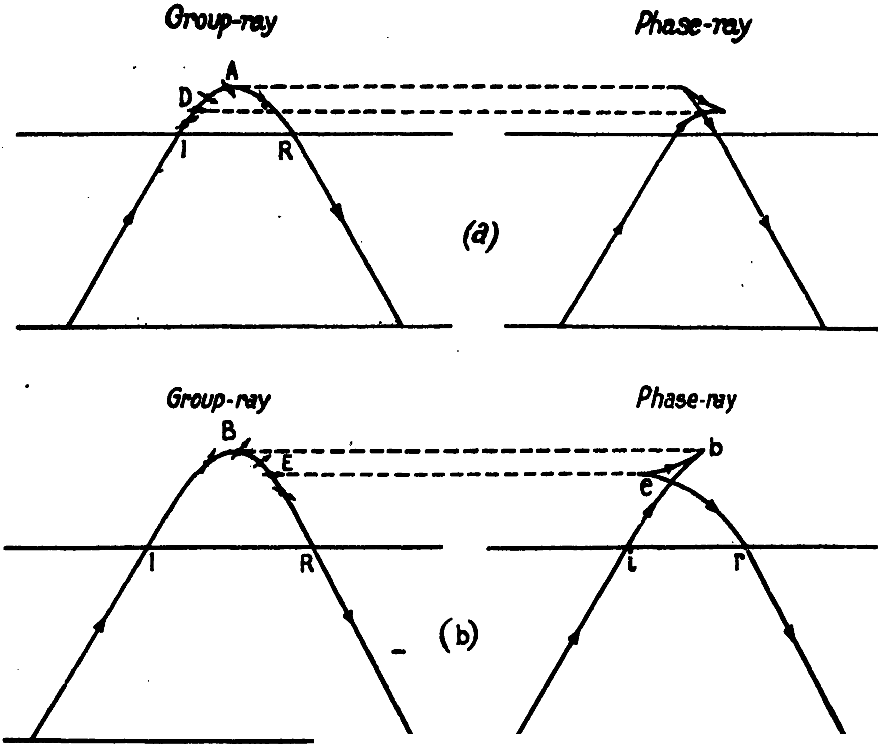


FIG. 27. Illustrating the path of the group-ray in relation to that of the phase-ray. (a) Extraordinary ray, and (b) ordinary ray. (After Booker.)

In this case the variation of q with N_e is given by the curves in Fig. 24 and the conditions $q=0$ and $dq/dN_e = \infty$ are satisfied at the same point. It may be mentioned that not infrequently the former condition is set forth as the necessary and sufficient condition of wave-reflection. But from what has been said above, it is obvious that the more general condition is $dq/dN_e = \infty$ [Sec. 2(a)].

A complication arises when the effect of damping is taken into account. The inclusion of the collisional term has the effect of making q complex with the result that the values of q corresponding to the upgoing and downcoming waves no longer attain equality at any point. In general, the reflection now becomes partial instead of total. If, however, the minimum difference between the two values of q is so small as to be negligible compared with the rate of change of q per $\lambda/2\pi$, total reflection may be considered to take place [31]. If the difference is of comparable magnitude to the rate of change of q per $\lambda/2\pi$, the reflection is only partial. If the damping is so large that the mimimum difference is very large compared to the rate

of change of q per $\lambda/2\pi$, no reflection takes place. The waves simply travel upwards and are absorbed.

As mentioned in the beginning of this section, the generalization of the magneto-ionic theory for the case of oblique propagation (in a horizontally stratified ionosphere) as described above is due to Booker. Booker has also applied his generalized theory to calculations of attenuation, lateral deviation, horizontal range and equivalent path of wave-packets incident at any angle to the vertical [31a]. The fundamental formula Eq. (45) being too complicated to be of practical use for this purpose has been re-stated. A cubic equation is obtained in which N_e is regarded as a function of q rather than q as a function N_e. 'Using this cubic equation, simple limiting curves between which the actual curves necessarily lie are deduced, thereby facilitating plotting in a particular case. Methods for mass-production of such propagation-curves are devised but not applied on large scale.

Formulas for attenuation, lateral deviation, horizontal range and equivalent paths of magneto-ionically split wave-packets are derived and illustrated graphically. The amount to which a wave-packet can be permanently deviated out of its plane of incidence by the earth's magnetic field can be extremely large, but probably not in cases of practical importance in commercial radio communication.'

Summary

We summarize below the results obtained above as it is only with their help that the results of experimental observations on the ionosphere can be properly interpreted.

The ionosphere is supposed to be horizontally stratified, the electron density increasing with height. The change in ionization density within a distance of one wavelength is assumed to be insignificant; the method of geometrical optics—the ray treatment—is thus applicable.

Taking first the simple theory, when no magnetic field is present, let a radio wave be incident on the boundary at an angle i. The condition for 'reflection' is given by

$$\sin^2 i = \mu^2 = 1 - \frac{4\pi N_e e^2}{mp^2}.$$

If the incidence is vertical, as is generally the case for study by the 'pulse' method (Sec. 4), the condition for reflection is

$$\mu^2 = 0$$

$$\text{i.e.,} \qquad \frac{4\pi N_e e^2}{mp^2} = 1.$$

If magnetic field is present, then the single incident wave will be split up into two waves—ordinary and extraordinary. Two cases may be distinguished according as the gyrofrequency He/mc of the electron, due to the steady magnetic field of the earth, is less or greater than p, the angular frequency of the wave.

(i) $He/mc < p$.—The condition for reflection is satisfied for three different values of N_e at three different heights. These are given by the equations

$$\frac{4\pi N_e e^2}{mp^2} = 1 - \frac{He}{mcp},\ 1,\ \text{or}\ 1 + \frac{He}{mcp}.$$

(ii) $He/mc > p$.—In this case there is, in general, two values of N_e at two different heights for $\mu^2 = 0$. These are given by

$$\frac{4\pi N_e e^2}{mp^2} = 1 \text{ and } 1 + \frac{He}{mcp}.$$

A remarkable fact about these conditions for reflection is that they are independent of the direction of the magnetic field.

When the direction of propagation coincides with the magnetic field, there are only two values for the first case ($He/mc < p$):

$$\frac{4\pi N_e e^2}{mp^2} = 1 \pm \frac{He}{mcp}$$

and one value for the second case ($He/mc > p$):

$$\frac{4\pi N_e e^2}{mp^2} = 1 + \frac{He}{mcp}.$$

These conditions hold strictly for the case of no collision between electrons and ions and neutral gas molecules or atoms.

The magnetically split waves are, on entering into the ionosphere, in general, elliptically polarized in opposite senses. The polarization ellipses are of the same eccentricity and are inclined at right angles to one another. In the northern hemisphere, the left-handed component (looking along the direction of propagation) is reflected at the point where $4\pi N_e e^2/mp^2 = 1$, and the right-handed component where $4\pi N_e e^2/mp^2 = 1 \mp He/mcp$. The state of polarization continually changes as the waves enter further and further into the ionosphere. Both the waves would be plane polarized at the level of electron density given by $4\pi N_e e^2/mp^2 = 1$. Actually, however, the right-handed component may not reach this region because its condition for reflection will be attained at a lower level were $4\pi N_e e^2/mp^2 = 1 - He/mcp$. It may also be noted in this connection that though there are two values of N_e for the reflection of the right-handed wave, it has usually little chance of reaching the higher level of greater electron density given by

$$\frac{4\pi N_e e^2}{mp^2} = 1 + \frac{He}{mcp}.$$

It is important to remember the change of the state of polarization on reflection of any one of the split waves. Taking the right-handed component, for instance, if one looks upward towards the ionosphere, he will observe the magnetic vector rotating in the right-handed sense both for the direct upgoing and for the reflected downcoming wave. From our definition, however, the upgoing wave is polarized in the right-handed

sense and the downcoming wave, looking down from above along the direction of propagation, is polarized in the left-handed sense.

The actual state of polarization of the downcoming wave on emergence from the ionosphere, as observed near the ground, depends upon the conditions near the boundary, that is, on the electron content and the angle between the wave-normal and the magnetic field.

Collision of the electrons affects the dispersion and the state of polarization. There is a critical collisional frequency ν_c given by

$$\nu_c = p_T^2/2p_L = \frac{H_T^2}{2H_L} \cdot \frac{e}{mc}.$$

If $\nu < \nu_c$, the propagation is of the quasi-transverse type. If $\nu > \nu_c$, the propagation is of the quasi-longitudinal type.

The effect of collision on the state of polarization of the downcoming wave is to rotate the polarization ellipse.

3. PROPAGATION OF ELECTROMAGNETIC WAVES IN AN IONIZED ATMOSPHERE (WAVE TREATMENT)

The wave propagation phenomenon as discussed in the preceding section is on the assumption that the extent of any region of inhomogeneity in the ionized medium is large compared with the wavelength. This, in other words, means that the ionization density (that is the refractive index) is supposed to vary slowly with reference to the length of the wave in the medium. In actual propagation in the ionospheric regions cases arise, however, where these conditions do not hold. In such cases recourse has to be had to the methods of wave optics instead of those of geometrical optics. We give here some instances of the failure of the geometrical optics method, i.e., of the ray treatment.

In the ionosphere there are regions in which the ionization density does not vary smoothly—the ionization being of a 'patchy' nature. In such a medium there is scattering both in forward and backward directions of the incident wave and the ray treatment fails when the size of the patches becomes comparable with the wavelength. In fact, it is found that the penetration of the wave into the region is controlled not by the maximum electron density but by the condition that the size of the patches is about $1/(4\pi)$ times the wavelength. (These points will be further discussed in Sec. 13(c).)

A second instance is provided by the case of partial reflection. It is often found that when the frequency of a wave incident on an ionospheric layer approaches the critical frequency the wave is only partially reflected, a fraction of it penetrating into the layer. According to the ray theory there is no room for such partial reflection. The incident wave will be either totally reflected or will penetrate according as the wave frequency is less or greater than the critical frequency. (We are considering, for simplicity, the case of no magnetic field and no collision.) Yet, according

to observations, such partial reflection and partial penetration phenomena are quite common. (See Sec. 4, Fig. 33.) The reason for this partial penetration is as follows:

From our simple consideration we have found that 'reflection' occurs when the *group velocity* becomes zero. But, we have, *group velocity* × *phase velocity* $= c^2$. Hence, as the critical frequency is approached, the group velocity tends to zero value and the phase velocity, and along with it the wavelength, tends to an infinitely large value. The dimensions of any inhomogeneity in the ionized medium become small compared with the wavelength and the ray treatment becomes inapplicable.

It is interesting to recall in this connection similar failure of the ray theory and the necessity of invoking the aid of the wave theory in optics. If a ray of light is incident on a thin film of refractive index smaller than that of the outside medium—the thickness of the film being comparable with the wavelength of light—then it is found that the ray will penetrate into the film even when it is incident at an angle greater than the critical angle as determined by the principles of geometrical optics.

As a third instance of failure of the ray treatment we consider the case of reflection of long- and very-long-waves from the bottom of an ionospheric layer. [Waves of length between 1 and 10 km. (frequencies between 300 and 30 kc/s) are called long-waves; waves of length greater than 10 km. (frequencies less than 30 kc/s) are called very-long-waves.] In Fig. 3 let the boundary AB be in the region below the E-layer. The ionization density in this region is known to increase from a negligible value to one sufficient for 'reflection' (by the bending process) of medium waves within the height of a few kilometres. Hence for long- and very-long-waves the complex refractive index (taking collision into account) changes by a significant amount within one wavelength and reflection takes place even though the refractive index has not fallen to zero value. Such waves may therefore be regarded as returning from the ionized layer not by a process of gradual bending (as shown in Fig. 3) but by reflection from a sharp boundary. Further, for the lowest frequencies the two magneto-ionic components are not propagated independently and the methods of geometrical optics again become inapplicable.

The general problem of the propagation of electromagnetic waves through the ionosphere, traversed by the terrestrial magnetic field, has not yet been exactly solved. This is because firstly, the equation of electromagnetic wave propagation has not been rigorously developed, and secondly, the boundary conditions (e.g., disposal of the antenna, condition of the ground, physical condition of the ionosphere) are not easy to be properly formulated in the solutions of the equation. However, for some special cases of ionization distribution and specified directions of the magnetic field (or, when the magnetic field is absent) equations have been developed and solutions obtained.

As an illustration let us first consider the case of partial reflection from an ionized layer of finite thickness which, according to the ray theory, acts as a

barrier to waves of frequency less than the critical frequency. The problem may be treated in a manner quite analogous to the treatment of crossing of potential barriers by fast particles in quantum mechanics (e.g., emission of α-particles from atomic nuclei) [33]. This was first done by Saha and Rai [34] who assumed for the purpose of calculation an electron barrier with an isosceles triangle profile. The treatment was later extended by Deb [35] to electron barriers of parabolic profile. This distribution approximates to the actual electron density distribution near the maximum of an ionized layer. Deb has determined the value of the reflection coefficient near the critical frequency following Kemble's analysis on the determination of transparency of a potential barrier of parabolic form by the so-called B.K.W. method in quantum mechanics [36]. According to Deb the transmission coefficient T for a wave of frequency f is given by

$$T = \frac{1}{1+\exp(2K)},$$

where $K = 4(\pi H/c)^{\frac{1}{2}} . \Delta f/f_c$, H—the 'scale height' of the region, c—velocity of light, f_c—critical frequency, $\Delta f = f - f_c$. It is evident from the above relation that for the true penetration frequency ($\Delta f = 0$) the value of the transmission coefficient T is 0·5. The reflection coefficient ρ ($=1-T$) has also the same value. ρ varies rapidly on either side of this frequency, attaining the value of 1 when Δf is positive and falling to zero when Δf is negative. These results explain the phenomena of partial reflection from and partial penetration into an ionized layer by waves of frequency close to the critical frequency.

Transmission properties of a parabolic layer of limited thickness ($\pm \Delta h_m$ from the plane of maximum density) when magnetic field, as also collision, are present have been studied by Rydbeck [37]. He has deduced the appropriate wave equation and has obtained exact solutions of the equation for two special cases of propagation—along and at right angles to the magnetic field. An interesting deduction is made in regard to the virtual height of reflection. It is found that for frequencies much below the critical frequency, the virtual height is about the same as that deduced from ordinary considerations. But, if the wave frequency is close to or greater than the critical frequency, then the equivalent height, instead of steadily increasing towards infinity, diminishes after reaching a maximum. However, no reflected wave can be detected, because the absorption increases very greatly near the critical frequency. Wave propagation between the earth and a concentric ionized layer with parabolic gradient (as also ionization varying as the square of the height) has also been considered by Rydbeck. As expected, it is found that the return of long waves from the ionized layer may be regarded as reflection from a sharp boundary.

An interesting feature of long distance propagation of long- and very-long-waves may be noted in this connection. Over long distances the propagation is in the space confined between the two concentric reflecting surfaces—the earth and the ionospheric layer. The propagation

phenomenon is then similar to that in a wave-guide, in which only a limited number of 'modes' (the number approximately equal to the number of waves which may be contained within twice the width of the guide) can be propagated without appreciable attenuation. Each of the modes in such wave propagation may be regarded as made up of two component progressive plane waves with their normals making equal and opposite angles with the horizontal. According to ray treatment we also have reflected rays of different orders. But these reflected rays are not to be confused with the two progressive plane waves mentioned above. It is interesting to note in this connection that round-the-world echoes have been recorded with very-long-waves (18 kc/s) propagated presumably by the wave-guide process [37*a*].

Amongst other works on wave treatment, making simplifying assumptions, mention may be made of those of Hartree, of Wilkes, of Rawer and of Stanley [37*b*]. Hartree has attacked the problem assuming that the ionization gradient is linear and the barrier has a 'triangular' profile [18]. Assuming similar ionization gradient Wilkes has deduced and solved the wave equation for the case of reflection of very-long-waves (plane) when the magnetic field is vertical [38, 39]. Rawer has worked with Epstein layer and calculated the transmission and reflection coefficients [40].

General wave equations have been deduced and discussed by Bose [41]. The wave equations in their complete form have been deduced by Saha and Banerjea [42]. They are as follows:

$$\nabla^2 E_z + \frac{p^2}{c^2}\left(1 - \frac{p_0{}^2}{(p^2 - c\nu p)}\right)E_z = 0$$

$$\nabla^2(E_y \pm iE_z) + \frac{p^2}{c^2}\left(1 - \frac{p_0{}^2}{p^2 - i\nu p \mp pp_z{}'}\right)(E_z \pm iE_y) = 0.$$

Here, the axes chosen are: x, in the direction of the magnetic field H; y, perpendicular to the magnetic meridian; and z, in the magnetic meridian perpendicular to H. Z is thus in general not in a vertical direction.

It has also been shown by Banerjea [43] from wave considerations that the single condition $\Psi = \frac{1}{2}\pi$, where Ψ is the angle between the electric and the displacement vectors, is equivalent to the various conditions for reflection as obtained by the different workers (see Sec. 2). Further, Saha, Banerjea and Guha [44] have deduced from wave considerations, expressions for refractive indices of the ordinary and the extraordinary rays and of polarization and absorption for any direction of propagation. These expressions are identical with those obtained from the Appleton-Hartree treatment.

4. RADIO SOUNDING OF THE IONOSPHERE

(*a*) Introduction

Of the various radio methods that have been developed for exploring the ionospheric regions, the one that is now almost universally employed

is the so-called 'pulse method' first introduced by Breit and Tuve [14]. In what follows we will first list the various ionospheric quantities that can be measured by the radio methods of exploration—in particular, by the pulse method [45]. Brief references will then be made to the principles underlying the different methods. Details of a typical equipment for measurement by the 'pulse' method will then follow.

The ionospheric quantities that can be measured immediately by radio sounding are the following:—

(*a*) The time taken by the radio signal to travel up to and down from the ionosphere.

(*b*) The intensity of the reflected signal.

(*c*) The state of polarization of the received signal.

Observation (*a*) yields the so-called equivalent height (h') of reflection (*vide infra*). From (*b*) the overall reflection coefficient ρ can be determined. The observed results in (*c*) can be interpreted in terms of the magneto-ionic theory. Further, the pulse technique enables one to measure the maximum electron number density of an ionospheric 'layer'.

Angle of incidence method.—In this method the angle of incidence of the downcoming ray (Fig. 28) is measured [46, 47]. Knowing the distance between the transmitting and the receiving stations, simple triangulation gives the equivalent height AG of the reflecting layer.

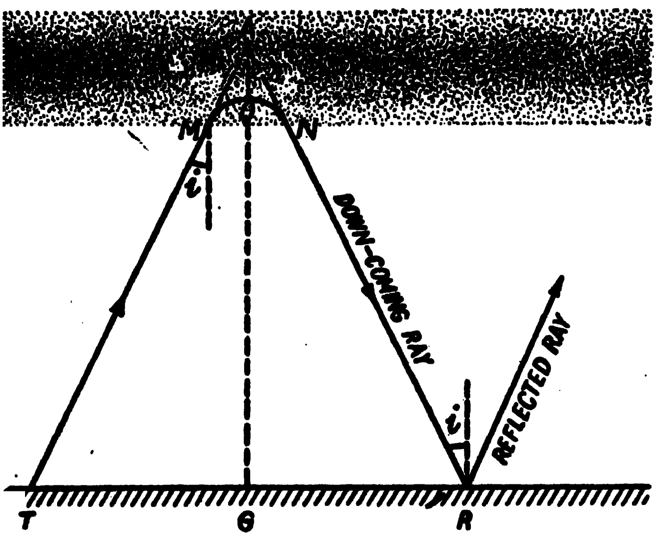


FIG. 28. Illustrating the principle of measuring the equivalent height of reflection from the angle of incidence of the downcoming wave.

Wavelength Change method.—The principle of this method [12] is as follows: If the waves sent out from a transmitter reach the receiver by two

paths—one along the ground and the other reflected from the ionospheric layer, then the resultant intensity will be a maximum if the two meet in the same phase, and a minimum if they meet in opposite phases. Now, if the wavelength of the transmitter is changed continually over a small range, the signal intensity goes through a number of maxima and minima and from a record of this number the equivalent height may be computed.

Hollingworth method.—This method is particularly suited for the determination of the equivalent height of reflection of long- and very-long-waves [48]. It consists in the measurement of the amplitude of the vertical electric field at the ground (or, in an aeroplane flying at a constant height) at different distances from the transmitter. The amplitude is found to fluctuate in a manner characteristic of an interference pattern formed by the superposition of the ground wave and the reflected wave. From the structure of the pattern the apparent height of reflection as also the reflection coefficient are easily determined.

Pulse method.—In the pulse method [14] the time taken by a series of radio frequency 'pulses' of short duration to travel up to the ionized region and return to earth is measured (Fig. 29). Knowing this time, the equivalent height of the reflecting region is easily calculated. Further, the amplitude of the reflected pulses is a measure of the reflection coefficient of the ionospheric layer from which the pulses are returned. The method is also called 'group-retardation method' because, as will be seen from the figure, the signal in its passage through the ionospheric region travels with the velocity appropriate for a wave-group.

Referring to Fig. 28 we note that the radio signal, in order to reach the receiver after reflection at the ionospheric region, has to travel along the path $TMNR$; of this, the lengths TM and NR are traversed in air ($\mu = 1$) with the velocity of light in free space. The length MON, being in the ionized region ($\mu < 1$), is traversed with a velocity less than that in air—the group-velocity of the waves in the ionized region. The time taken by the signal to traverse the path $TMNR$ is, therefore, given by

$$\int_{TMNR} \frac{ds}{v},$$

where ds is an element of path and v is the velocity of the wave-group at any point in its path. It is this time that is measured by the group-retardation method. It will be shown later [Sec. 4(e)] that this time is the same as that taken by a signal to traverse the entire path TAR (Fig. 28) with the velocity of light in free space. The path TAR is, therefore, called the *equivalent path* P'. Thus

$$P' = c\int_{TMNR} \frac{ds}{v}.$$

The height computed by any of the methods mentioned above gives the height (AG) of the vertex of the equivalent triangle, and is called the *equivalent height*. The equivalent height is thus always greater than the true height OG of the point of reflection.

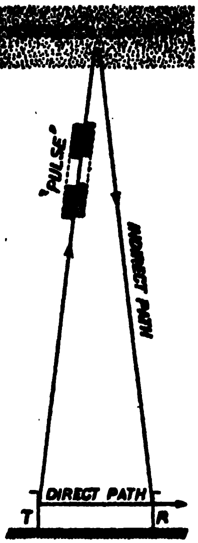


FIG. 29. Illustrating sounding the ionosphere by radio-wave 'pulses'.

To determine the maximum electron number density N_{max} of an ionized layer exploring waves of gradually increasing frequency are sent vertically upwards to meet the layer. Reflections will occur up to the frequency for which $4\pi N_e \epsilon^2/mp^2 = 1$. If the frequency is increased still further the waves will penetrate the layer and there will be no further reflection. The limiting frequency at which the reflections—the so-called 'echoes'—just begin to disappear, is called the critical penetration frequency of the layer. If $f_c (= p_c/2\pi)$ denotes this critical frequency, then

$$\frac{4\pi N_e \epsilon^2}{m(2\pi f_c)^2} = 1$$

or $$N_e = N_{max} = \frac{\pi m}{\epsilon^2} f_c^2 = 1{\cdot}24 \times 10^4 f_c^2, \quad \ldots \quad (46)$$

where f_c is expressed in megacycles per second. The determination of f_c, therefore, enables one to compute the maximum electron number density of the ionospheric layer penetrated.

The remarks made above regarding the true and equivalent heights of reflection and the computation of the maximum electron number density from the critical penetration frequency are applicable only when the ionized region is not traversed by any external magnetic field. When a magnetic field is present, the wave is split up into two components and the equivalent heights of reflection and the penetration frequencies of the split components—the ordinary and the extraordinary—have to be distinguished from each other.

Assuming that the echoes corresponding to the ordinary and the extraordinary waves have been identified, the maximum electron number density can be calculated by noting the critical penetration frequencies of these components.

Thus for the ordinary wave, if the critical penetration frequency be f_o, then

$$N_{max} = \frac{\pi m}{\epsilon^2} . f_o^2. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (47)$$

For the extraordinary wave two cases have to be distinguished. If the exploring frequency is greater than the gyrofrequency $(p > p_H)$, then, as seen in Sec. 2(c), the extraordinary ray may be reflected from two distinct regions of electron densities given by

$$p_0^2 = p^2 - pp_H \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (27)$$

and $$p_0^2 = p^2 + pp_H . \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (29)$$

Since the electron density increases with height, the region defined by (27) is at a lower level than that defined by (29). It is only under very special circumstances that the wave will have appreciable energy left after partial reflection at the level (27), to travel up to and undergo reflection at the level (29). Generally, most of the energy will be reflected at the lower level. If we denote the critical frequency for this layer by $f_x (= p_x/2\pi)$, then from condition (27)

$$\frac{4\pi N_{\max} e^2}{m} = 4\pi^2 f_x^2 - 4\pi^2 f_x f_H$$

or

$$N_{\max} = \frac{\pi m}{e^2}(f_x^2 - f_x f_H). \quad \ldots \quad \ldots \quad \ldots \quad (48)$$

If the exploring frequency is less than the gyrofrequency, then the extraordinary component is reflected from the level given by $p_0^2 = p^2 + pp_H$, from which, we get

$$N_{\max} = \frac{\pi m}{e^2}(f_x^2 + f_x f_H). \quad \ldots \quad \ldots \quad \ldots \quad (49)$$

There may also be a reflected ray (extraordinary) according to the fourth condition of reflection of the wave

$$\frac{p_0^2}{p^2} = \frac{p^2 - p_H^2}{p^2 - p_L^2},$$

as explained in Sec. 2(*c*). If f_x is the corresponding penetration frequency, then,

$$N_{\max} = \frac{\pi m f_x^2}{e^2} \cdot \frac{f_x^2 - f_H^2}{f_x^2 - f_L^2}, \quad \ldots \quad \ldots \quad \ldots \quad (50)$$

where

$$f_H = \frac{2\pi H e}{mc} \text{ and } f_L = \frac{2\pi H_L e}{mc}.$$

This condition will be further discussed in Sec. 4(*f*), in connection with the inclusion or otherwise of the Lorentz polarization term in the magneto-ionic formula.

(*b*) The 'pulse' equipment

(i) *Introduction.*—As already mentioned, most of the ionospheric explorations are carried out with equipments working on the group-retardation principle of Breit and Tuve. Such an equipment consists essentially of the following components:

(1) A transmitter generating radio frequency 'pulses' of short duration 50–150 microseconds. The wave frequency of the 'pulse' is capable of being varied continuously over a wide range, say 0·5–20 Mc/s. The pulses have peak power of several hundred watts and are repeated at some convenient frequency, 50 or 100 per sec.

(2) A broad-band communication-type receiver which can accept the band width of the pulses used. The tuning of the receiver is variable

continuously over the entire range of frequencies covered by the transmitter frequency. The receiver is also so designed as not to be paralyzed when placed close to the transmitter.

(3) Broad-band antennas, one for transmitting and one for receiving, such as vertical rhombics, or inverted V's, having good vertical directivity.

(4) Cathode-ray oscilloscope for displaying the received echoes.

We will now describe in some detail the method of recording the echo-pattern as observed on the oscilloscope screen. It is only from a proper interpretation of this record that the ionospheric characteristics are obtained.

The received echoes produce sharp integrated signals at the output of the receiver, which is connected to the Y-deflecting plates of the oscilloscope. The X-plates are connected to a linear time base, the sweep frequency of which is made equal to the pulse repetition frequency. By suitably synchronizing or locking the time base with the recurrent received pulse signals (with some relative phase shift so that the rising point of the direct ground pulse is clearly recognized) a steady pattern of the direct signals and echoes appears on the oscilloscope screen. The distance between the direct signal and the echo depends upon the time-retardation. If t_s is this delay time in microseconds, then the equivalent height h' of reflection is equal to $0{\cdot}15\ t_s$ kilometres. Fig. 30 shows a typical simple echo-pattern.

FIG. 30. Typical echo-pattern on the cathode-ray oscillograph screen of 'pulse' equipment. The signal on the extreme left is due to the direct ground ray.

The time base is calibrated and h' can be read off directly. (It is to be remembered that the time interval between the direct ground pulse and the echo corresponds to twice the equivalent height.)

For experimental determination of the maximum electron number density of a reflecting layer, the corresponding critical frequency f_0 is determined by gradually increasing the frequency of the exploring waves and noting the frequency at which the echoes just disappear. If the reflecting layer has a sharp boundary, as in the case of E-layer, then there is little observable increase of equivalent height with increase of frequency. If, however, the layer is diffuse as in the case of the F-layer, then as the critical frequency is approached, the equivalent height increases enormously, due to large time-retardation of the exploring pulses. In the latter case magneto-ionic double refraction is often observed, so that there are two echoes, one ordinary and another extraordinary. In determining the critical frequency it is helpful to plot the equivalent height (or rather the equivalent path P') against the exploring wave-frequency. These curves—the so-called $(P'-f)$ curves—are extremely important in ionospheric investigation, as, besides giving the maximum electron number density, they provide information regarding various other characteristics of the ionosphere.

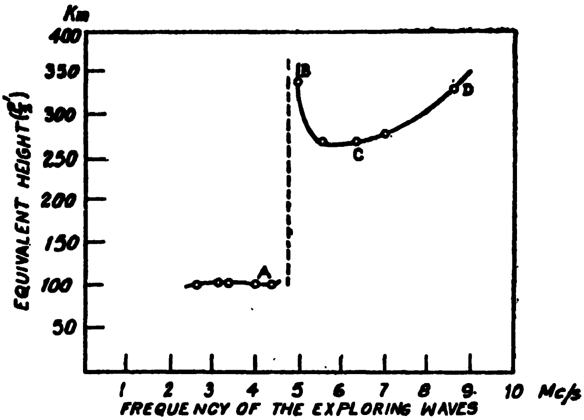


FIG. 31. Variation of equivalent height of reflection from E and F regions with change of frequency of the exploring waves. A—reflections from E region; BCD—reflections from F region. Note the abnormal increase of the equivalent height ($P'/2$) of reflection from the F region near the penetration frequency of the E region.

Fig. 31 shows an ($h'-f$) plot in which the exploring wave-frequency has been increased by small steps. The circles represent the echoes received. It will be noticed that at 4·6 Mc/s., echoes suddenly cease from the 100 km. equivalent height. The maximum electron density of the reflecting layer is then given by (Eq. 46)

$$N_{\max} = 1{\cdot}24 \times 10^4 \times (4{\cdot}6)^2.$$
$$= 2{\cdot}5 \times 10^5 \text{ electrons per cm.}^3$$

On further increasing the frequency, the echoes reappear from much greater equivalent height. These are due to reflections from Region F. It will be noticed that the equivalent height at B is very large. This is because

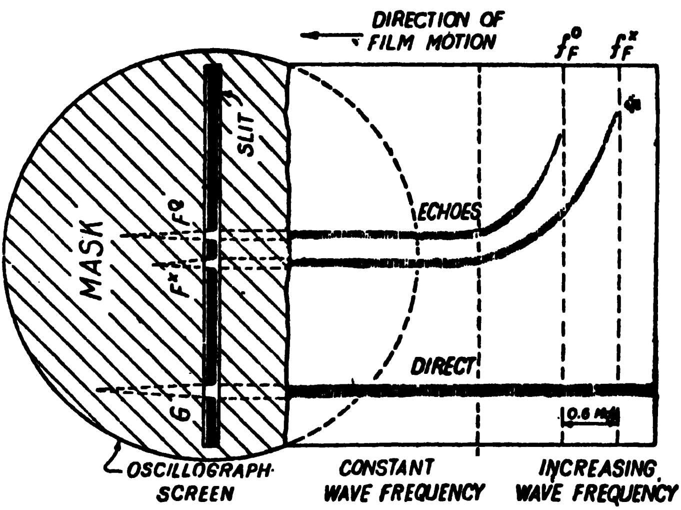


FIG. 32. Illustrating the method of continuous recording of ionospheric echoes. The shaded region is part of the mask covering the whole oscillograph screen except the time-base (thick black line). Any pulse appearing on the time-base produces a discontinuity which leaves its trace on the moving film.

the retardation of the exploring wave group by Region E near its critical penetration frequency is very large.

In modern equipments, the exploring wave frequency is varied continuously and the $(P'-f)$ curve is photographed directly on film. The method for doing this is as follows:

The oscillograph screen is masked in the manner shown in Fig. 32 and a box camera is used in conjunction with specially fast bromide paper or film. The film is moved slowly in a direction at right angles to the time-base by a motor and the breaks in the time-base due to the ground pulse and the echoes form traces on the film. There is a switching device for the motor in the camera unit which, when operated in one direction, automatically opens the lens shutter. Fig. 33 is a reproduction of an actual $(P'-f)$ record.

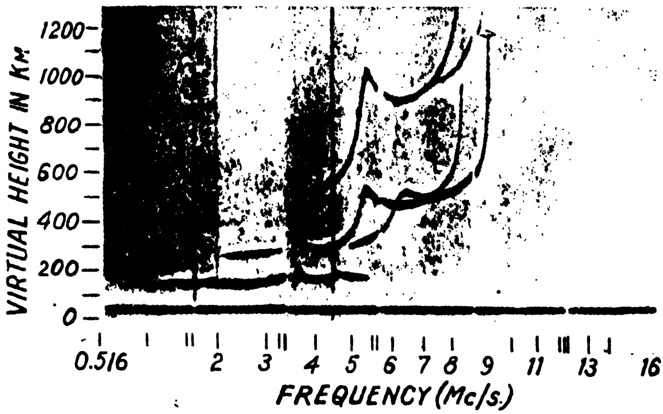


FIG. 33. Typical $(P'-f)$ curves obtained with an automatic ionospheric equipment (Booker and Seaton) [93]. Watheroo Magnetic Observatory, December 18, 1939, 16h. 15m. local time (120° East meridian).

From what has been said above it is evident that with the pulse technique one can determine the following two basic relations:—

(1) The manner in which the equivalent height of reflection (h') varies with time (t) and with wave frequency (f), i.e., the nature of the function, $h' = \phi_1(f, t)$,

(2) The manner in which the reflection coefficient (ρ) varies with time (t) and with wave frequency (f), i.e., nature of the function, $\rho = \phi_2(f, t)$.

However, experimentally it is simpler to deal with only one of the independent variables, the other being regarded as constant. Thus, in practice, we can obtain the following four relations:—

$$h' = [\phi_3(t)]_{f\,\text{const.}}; \quad h' = [\phi_4(f)]_{t\,\text{const.}}; \quad \rho = [\phi_5(t)]_{f\,\text{const.}}$$

$$\text{and,} \quad \rho = [\phi_6(f)]_{t\,\text{const.}}$$

(ii) *Automatic multifrequency ionospheric recorder.*—As the ionospheric characteristics are subject to rapid changes, it is necessary that the complete sweep of the frequency range is made in as short a time as possible. All modern ionospheric equipments have been developed with this end in view. Further, the modern equipments register the complete ($P'-f$) sweep automatically the transmitter and receiver being interlocked and always in tune [49]. Panoramic recorders have also been developed which permit the complete $P'-f$ curve to be delineated on the oscillograph screen. The different automatic equipments developed in the various important ionospheric laboratories of the world differ necessarily in details of construction—the frequency interlocks, the time for a complete sweep, etc. In what follows we will describe the principal features of an automatic ionospheric recorder as is in daily use at the ionosphere field station of the University of Calcutta, at Haringhatta, 28 miles (20° east of north) from Calcutta.

The apparatus is designed and built by the Commonwealth Scientific and Industrial Research Organization, Australia [50]. It makes a complete $P'-f$ record in 1 min. 55 secs. over a frequency range 1–13 Mc/s. This record is obtained on a standard 35 mm. film and is repeated every ten minutes, i.e., 6 records are obtained in one hour. The number of records can, however, be increased to as many as 30 per hour.

The cabinet, housing the complete apparatus consisting of transmitter, receiver, variable frequency tuner, recorder, aerial tuning system and the power units, mounted in their chassis is shown in Fig. 34 (Plate I).

The various components are arranged in the two cubicles, upper and lower. The upper part contains the transmitting valves with the modulator, tank-circuits, switches, receiver tuning units and receiver r.f. and converter valves. The lower part contains the receiver i.f. stages, detector, pulse circuits, circuits for the recorder and also the power units for all circuits—mounted on 3 separate chassis which can be conveniently slided in or taken out.

The main characteristics of this type of automatic apparatus, as distinguished from the manually operated ones, are as follows: (1) The transmitter and receiver tuning systems are always kept tuned to each other, (2) the oscillator frequency is slowly varied automatically over the entire range, and (3) the operation is repeated automatically at predetermined intervals. The first two are accomplished by controlling the movements of the transmitting and receiving tuning condensers mechanically by means of cams carried on a common cam-shaft supported through self-aligning ball races. The cam rotates slowly, being coupled to a motor through worm reduction gear-box. For controlling the programme of operation by repeating the performance at selected intervals, a timing shaft is coupled to the same motor after further reduction of speed.

The entire frequency range 1–13 Mc/s. has necessarily to be covered by more than one coil-condenser unit. In the apparatus described, 4 sets

of coil-condensers are used for transmitter tuning and another 4 sets for receiver tuning. For switching over from one unit to another, mercury-in-glass switches are used. We now proceed to describe the main constructional details of the components.

Transmitter.—Two 250-watt transmitting valves in push-pull disposition placed in the upper cubicle and drawing power from the power pack (housed in the lower cubicle) are modulated at the grid by square-topped 400 volts negative pulses of 150 μ sec. duration. The oscillator tank circuits are four in number corresponding to the four bands, 1–2·1, 2·1–4·4, 4·4–7·6 and 7·6–13 Mc/s. They are housed in the upper cubicle along with the mercury-in-glass switches connecting the tank circuits to the oscillator valves. The power from the transmitter is fed by transmission lines to a vertical Δ-type aerial.

Modulator.—The modulating pulses are generated by a thyratron, controlled by primary pulses from a multi-vibrator operated by 50 cycles a.c. mains. The thyratron is placed on the same chassis as the transmitting valves.

Receiver.—The receiver is of superheterodyne type with 3 i.f. stages. The tuning range of the receiver is divided into 4 bands corresponding to the 4 bands of the transmitter. Each tuning unit has its own r.f. and mixer stages and is interlocked with the corresponding band of the transmitter. These tuning units form part of the variable frequency tuner housed in the upper cubicle. The i.f. and detector stages of the receiver are placed in the lower cubicle. The selectivity of the receiver is adjusted to accept the band-width of the pulses without serious loss in sensitivity. The circuit dispositions are also so planned that the receiver recovers quickly from the paralyzing effect of the direct transmitter pulse.

Recording System.—The recorder is a cathode ray oscilloscope with a film camera. The output of the receiver (which is always maintained in tune with the transmitter) is applied to the vertical deflecting plates of the oscilloscope. The echo pattern produced on the oscilloscope screen is photographed continuously on a moving film by a camera mounted on a swinging bracket situated in the lower cubicle. The film is standard 35 mm., drawn at the rate of 0·8 inch per minute, and 18 feet of film is consumed daily. The time base of the oscilloscope is generated by a circuit driven by the main multivibrator pulses. Height marker pulses are generated by a 3000 c/sec. oscillator also controlled by the primary multivibrator pulses. The record for a complete 'sounding' of the ionosphere is produced on a film length of about 1¼ in. An enlarger or projector apparatus is required for reading the film. With the calibrating marks printed on each record (at 50 km. intervals for height and 0·5 Mc/s for frequency) the heights and the penetration frequencies can be read off to accuracies of $\pm$ 10 km. and $\pm$ 0·1 Mc/s. respectively.

Power-Supply.—Power for the various circuits of the receiver and recorder are taken from the power supply unit housed in the lower cubicle.

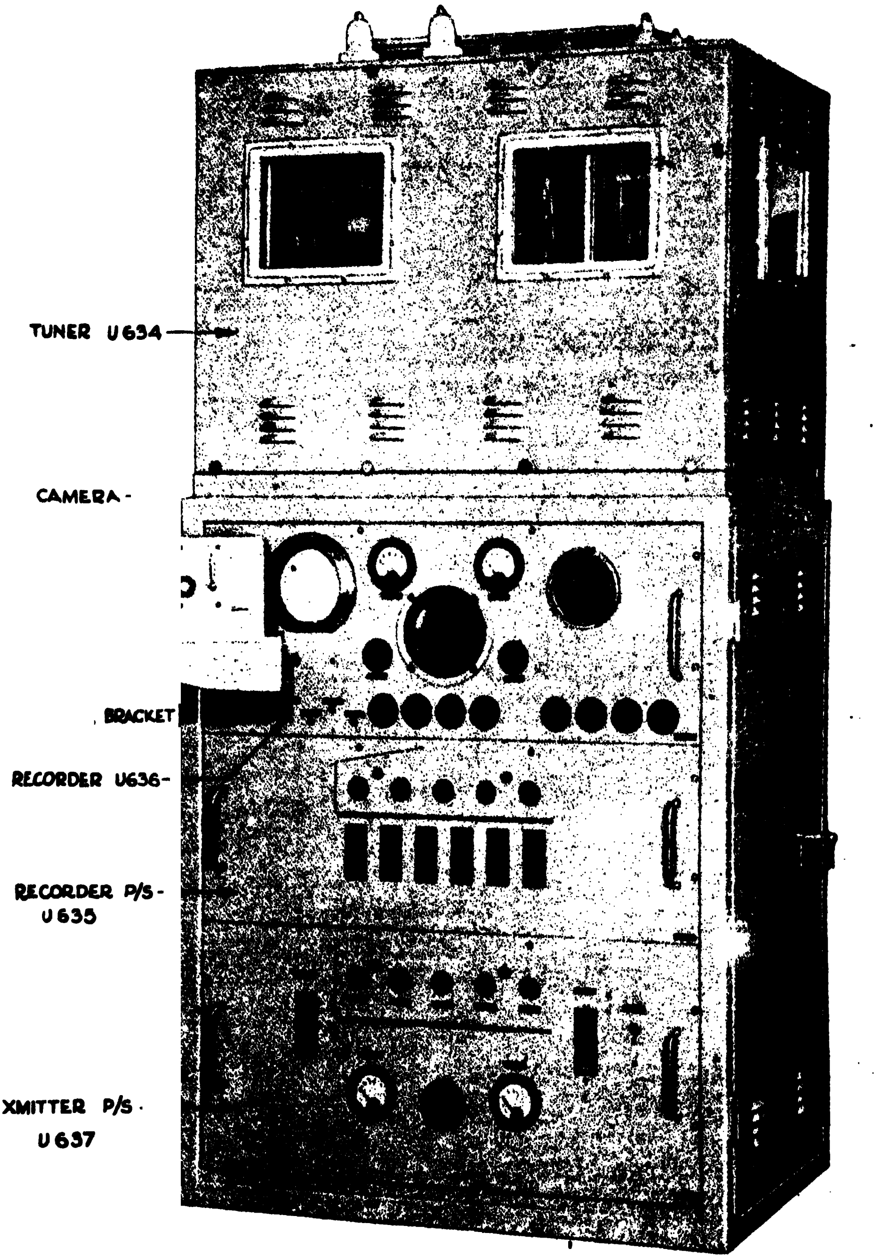


FIG. 34. Front panel of an automatic ionospheric equipment designed and built by the Commonwealth Scientific and Industrial Research Organization, Australia.

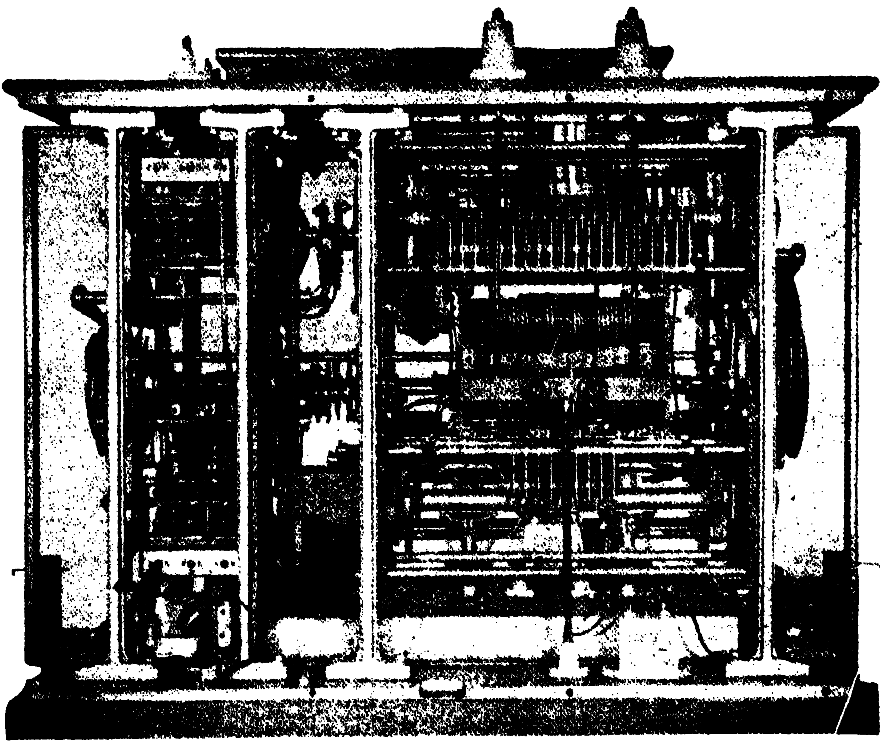


FIG. 35. Front view of automatic tuner of the ionosphere apparatus shown in Fig. 34 (Plate I).

The variable frequency tuner.—This is the most important part of the automatic apparatus and is housed in the upper cubicle. It consists of the mechanisms for controlling the tunings of the transmitter and the receiver, and for the repetition of the frequency sweep at predetermined intervals.

Both the transmitter and the receiver variable condensers are specially designed and are driven mechanically, through cams and rocker arms as already mentioned. Special attention has been paid in the design of the cam shaft to ensure that, once these cam contours have been correctly cut, the frequency law is maintained and the transmitter and receiver remain in tune.

The determination of the cam contour is done experimentally by marking out first on an over-size cam-blank, successive positions at close intervals of the cam followers and then cutting the cam to the envelope of the curves obtained. Transmitter cams can be cut only at the final site with the transmitting aerials connected to the equipment. The receiver cam can be cut more easily if suitable facilities are available. Symmetrically around the cam shaft are placed, at one end, the four transmitting condensers, and at the other end the four receiving condensers. At each end there are four shafts carrying rocker arms. Control springs are fitted to all rocker arm shafts to ensure that the cam follower is always in contact with its cam and to restore the condenser after operation to its original setting. Fig. 35 (Plate II) shows the assembly.

As has already been said, switchings from band to band for both transmitter and receiver and the automatic on and off switching are done by cam gear operating mercury switches. These switches again are controlled by a further set of switches operated by cams on a shaft rotating once in ten minutes.

The equipment is provided with ample safety factors for continuous operation. Once started, it can switch itself off and on automatically and may be left unattended.

(iii) *Panoramic recorder.*—In recent years there has been a useful development of ionospheric recorder known as 'panoramic recorder'. In this type of recorder the panoramic principle as used in radar equipment is employed with certain modifications. A large cathode-ray tube with long afterglow is used and one obtains on the screen a standing picture of the reflections from the various ionospheric layers. Such a picture is obviously of great use as it makes it possible the study of rapid ionization changes in the different layers (e.g. during radio fade-outs and magnetic storms) as also the short-lived sporadic E-echoes.

The oscillographic recording arrangement is such that the ground pulse and the echoes appear as bright spots and the time base sweep is made to move parallel to itself as the frequency is increased. It is clear that if the frequency sweep is sufficiently quick, and the afterglow is sufficiently long, the P'—f trace will be observed directly on the screen as a bright line.

The special refinements of this type of recorder are thus: (1) Bright spot records, (2) fast frequency sweep (5–10 secs.) and (3) arrangement for

slow movement of the time base parallel to itself as the frequency is gradually increased.

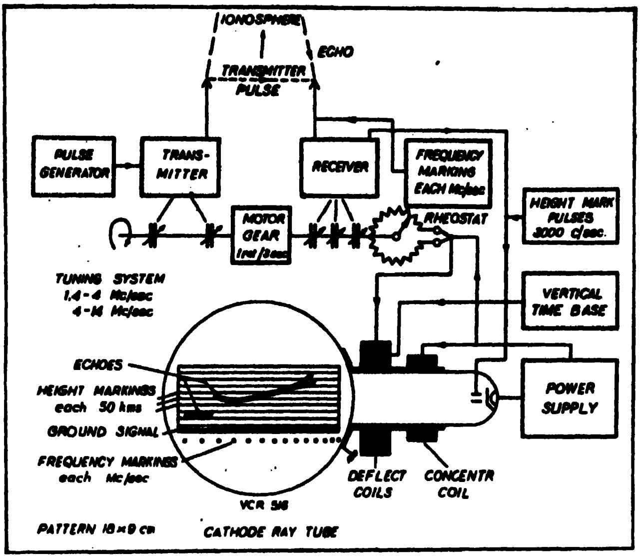


FIG. 36. Schematic diagram illustrating the principle of construction of panoramic ionosphere recorder. (After Stoffregen.)

Fig. 36 illustrates the principle of the construction of a panoramic recorder as designed and built by Stoffregen of the Norwegian Defence Research Establishment. The recorder is in use at several stations in Norway for study of ionospheric storms as occur during magnetic disturbances and auroral displays in high latitudes [51].

The receiver and the transmitter are tuned together mechanically by means of a correction wheel. An improved synchronous electrical method has also been developed for tuning together the transmitter and receiver. In this, the method of generating the transmitter signals is to mix the oscillator signal of the superheterodyne receiver with the signal of a second oscillator (pulse modulated). The frequency range of the apparatus is from 1·4–14 Mc/s. The cathode-ray oscillograph screen has an afterglow of about 10 secs. and the size of the panoramic picture is 18 cm. × 9 cm. Record for each picture takes 3–5 secs. to be completed. The vertical deflection line moves from left to right in synchronization with the rotation of the high frequency tuning system and thus produces the afterglow pattern. The speed of the recorder may be reduced considerably for ordinary routine observation.

Panoramic recorders have also been developed at other centres of ionospheric research. Some of these [52, 52*a*, 52*b*], as are in regular use at the time of writing (1950) are at: (1) Washington D.C. (38·7°N., 77·1°W.), U.S.A., (2) Kiruna (67·9°N., 20·4°E.), Sweden, (3) Johannesberg (26·2° S., 28·0° E.), South Africa, (4) Pennsylvania (41·5° N., 78° W.), U.S.A.

(*c*) Complex echo-patterns: Group-retardation and stratification splitting—Intensity of the magnetic field in the ionospheric regions

As already mentioned the echoes on the oscillograph screen are not always simple; very often they form a complex pattern. The complexity arises from various causes, the principal ones being complex structure of the ionosphere and magnetic double refraction of the waves travelling through it. In order to deduce the ionospheric characteristics from the (P'—f) curves traced out by these echo-patterns it is very necessary that the nature of the echoes be correctly interpreted.

Multiple echoes may be obtained on account of multiple reflections between the ground and the ionospheric regions. They are also obtained when the exploring wave is reflected partly by one and partly by another ionospheric region at a greater height. This happens when the wave-frequency is near the penetration frequency of the lower region. Complex echo-patterns due to the above two causes do not present any difficulty for their interpretation. The interpretation of the components of a split echo due to magnetic double refraction is, however, not always easy.

We recall that a wave on entering the ionosphere is split into two waves. They travel into the ionospheric regions with different velocities and reach different heights before reflection. Under suitable condition the relative retardations of the two waves, due to these two causes, might be sufficiently large to produce on the oscillograph screen two distinct echoes, one due to the ordinary and the other due to the extraordinary wave. Since there are two factors controlling the retardation, one cannot always predict which of the two split components—the ordinary or the extraordinary—will return earlier. We thus distinguish between two types of splitting, the group retardation splitting in which the difference in retardation between the components is due to the two components travelling with different velocities, and, stratification splitting in which the difference is due to the component waves being returned from different strata of the ionosphere.

Group-retardation splitting.—In Fig. 37, the wave in passing through the ionized region A is split into two components—ordinary and extraordinary—moving with different group velocities v_0 and v_e. If the exploring wave-frequency be greater than the gyrofrequency then $v_0 > v_e$. If there is an ionized region B above A of sufficient density, then the two split waves are reflected by this region and, of the two downcoming waves, the extraordinary will be more delayed if it is assumed that the two waves are reflected from nearly the same level of region B. This type of splitting, though of infrequent occurrence, is possible if there is large group-retardation in Region E and reflection of both the components takes place from nearly the same level of Region F. In this type of splitting, therefore, the long delay component is the extraordinary wave and is polarized right-handed in the northern hemisphere.

Stratification splitting.—This type of splitting is of more frequent occurrence and its origin may best be understood by referring to Figs. 12

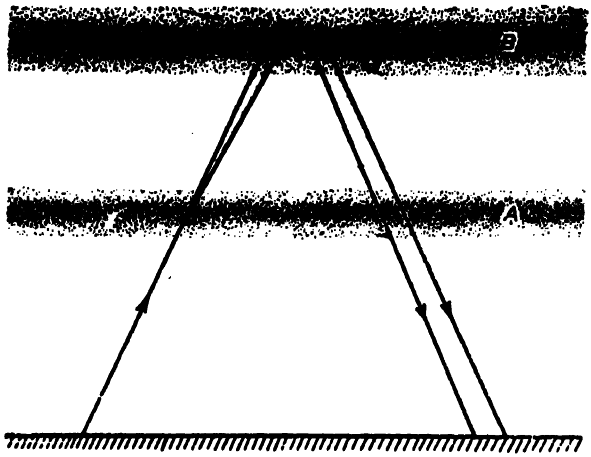


FIG. 37. Group-retardation splitting. The wave-group entering the lower region A is split up into two components—ordinary and extraordinary—which travel with different velocities. The two components are reflected from the upper region B.

and 13. In Fig. 13, for the case of wave-frequency less than gyrofrequency, the ionization densities necessary for reflection of the two components are given by $1+\alpha = 0$ and γ. In Fig. 12, for the case of wave-frequency greater than gyrofrequency there is one condition for ordinary wave $1+\alpha = 0$ and three conditions for extraordinary wave reflection, namely, $1+\alpha = \pm\gamma$ and the so-called fourth condition (Eq. (30)). Now imagine a wave-group to be propagated vertically in the ionosphere into regions of increasing electron density. It is obvious, taking the case of Fig. 12, that the two split components of the incident wave-group will be reflected from two different levels. Since the density increases from bottom upwards, the extraordinary will be reflected from the lowermost level $(1+\alpha = -\gamma)$ and the ordinary from a higher level $(1+\alpha = 0)$. In the echo-pattern for such a case the long delay component is the ordinary and the short delay component is the extraordinary. This is because the ordinary wave penetrates deeper into the ionized region than the extraordinary wave and is more retarded. This retardation more than balances the opposite retardation caused by the ordinary wave travelling faster.

Multiple splitting.—Besides such simple splitting which is of common occurrence, triple or even quadruple splitting is occasionally observed [53, 53*a*]. Such triple or quadruple splitting of the echoes is not to be thought of as due to triple or quadruple splitting of the wave entering the ionosphere. As remarked in the beginning, a single wave-group is split into two and only two components due to the influence of the earth's magnetic field. The multiple splitting of echoes is due to partial reflections of the extraordinary wave from three possible levels of reflection. The occurrence of such multiple

splitting may again be understood by reference to Fig. 12. We have already noted that the extraordinary wave has three reflection conditions each corresponding to a particular value of N_{max}. It is possible that the extraordinary wave-group instead of being totally reflected from the lowermost level (where the particular value of N_{max} satisfying the reflection condition is the least) is only partially reflected [Sec. 2(c)] from all the three levels. In such case there will be three split components as extraordinary and one as ordinary.

The polarization, the nomenclature and the relative intensities of the components are all related to the direction of propagation of the wave with respect to the magnetic field. The relations are shown in tabular form below.

TABLE I

Downward propagation in the northern hemisphere

Exploring wave frequency	Nature of split components	Polarization	Absorption
Less than gyrofrequency	Ordinary (short delay)	Left-handed	less absorbed
	Extraordinary (long delay)	right-handed	more absorbed
Greater than gyrofrequency	Ordinary (long delay)	left-handed	less absorbed
	Extraordinary (short delay)	right-handed	more absorbed

On account of larger absorption the extraordinary component may be altogether absent. Further, if the ionization gradient is sharp then the two components may not be resolvable in which case the composite echo is called a non-split echo.

Determination of the earth's magnetic field from ionospheric observations.—The phenomenon of stratification splitting may be utilized for estimating the earth's magnetic field in the ionosphere [53*b*, 54]. The experiment consists in observing the disappearance of the two split components (stratification splitting) for Region F reflection say, by gradually increasing the frequency. If the frequency employed is greater than the gyromagnetic frequency and if, as is usually the case, the mode of propagation for these frequencies is quasi-transverse (collisional frequency ν less than the critical collisional frequency ν_c) the maximum electron number density N_{max} is related to the critical frequencies by the equations:

$$N_{max} = \frac{\pi m}{e^2} f_o^2 \qquad \text{(ordinary)} \quad .. \quad (47)$$

and

$$N_{max} = \frac{\pi m}{e^2} (f_x^2 - f_x f_H) \quad \text{(extraordinary)}, \quad .. \quad (48)$$

where $f_H (= He/2\pi mc)$ is the gyromagnetic frequency, and f_o, f_x the critical frequencies for the ordinary and extraordinary components respectively. Equating these two expressions we have

$$f_o^2 = f_x^2 - f_x f_H$$

or

$$H = \frac{2\pi mc}{e} \cdot \frac{f_x^2 - f_o^2}{f_x}. \quad \quad .. \quad .. \quad .. \quad (51)$$

If the propagation is of the quasi-longitudinal type, i.e., if $\nu > \nu_c$, the two expressions for $N_{\max}$ are

$$N_{\max} = \frac{\pi m}{e^2}(f_o^2 + f_o f_H) \quad \text{(ordinary)}$$

and

$$N_{\max} = \frac{\pi m}{e^2}(f_x^2 - f_x f_H) \quad \text{(extraordinary)}.$$

Equating these two expressions we have

$$H = \frac{2\pi mc}{e}(f_x - f_o). \quad \quad .. \quad .. \quad .. \quad (52)$$

Eq. 51 or 52 can be utilized for determining the intensity of the earth's magnetic field in the ionospheric regions. Such determination was first made by Appleton, at Slough (55°N.) England [54]. The value of H was found to be 0·42 Gauss at the height of Region F, a value 10% less than the ground value. This agrees with the calculated value on the assumption that the earth is a uniformly magnetized sphere (Chap. VII, Sec. 1). The value of H in Region F has also been calculated from a large number of observations at Calcutta (23°N) [55], the mean value found being 0·36 Gauss. (The ground value of H at Calcutta is 0·43 Gauss).

An interesting fact which needs corroboration has been reported by Scott [56] from similar observations made at Clyde River, Baffin Islands (70·5 N.) Churchill, Canada. The intensity of the earth's magnetic field in Region F, as calculated from Eq. (51), was found to have a large seasonal and diurnal variation, the value being greater in summer than in winter, though the diurnal variation was larger in winter than in summer. The magnitude of the variation was found to be as much as 20%. It may be noted that in the observations at Calcutta referred to above, some significant seasonal variation in the value of H, as deduced from Eq. (51), had also been noticed. The magnitude, however, is smaller, not exceeding 10%.

(d) Structure of the ionosphere

As mentioned in the introduction and as shown in Fig. 1, the density of ionization in the ionosphere is not uniform throughout. The ionosphere is stratified into a number of regions or layers of maximum ionization

densities.* Four such regions, designated D, E, F_1 and F_2, are shown in the figure. Of these, F_2 and E are the more persistent ones. The intermediate region F_1 is observable only during daytime. It coalesces with F_2 at night forming a single region F (Fig. 38). Region D is produced only during daylight hours and is mainly an absorbing region particularly for waves of medium frequencies. The E and F_2 regions are also known as *Kennely-Heaviside* and *Appleton* layers respectively.

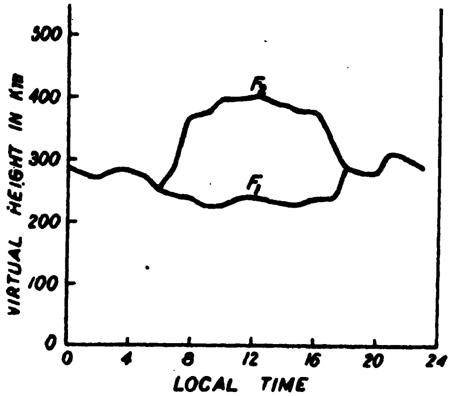


FIG. 38. Illustrating the day-time splitting of Region F into two Regions F_1 and F_2. The curves depict how the *virtual height of maximum ionization* varies throughout 24 hours (Canberra, December, 1947). (The virtual height of Region F_2 as illustrated is rather unusually high. The average height is round 300 km.)

The existence of the E, F_1 and F_2 stratifications in the ionosphere was discovered from a study of the variation of the equivalent height P' with the frequency f. Fig. 39 is prepared from a typical ($P'-f$) record obtained during daytime. It will be seen that there are discontinuities in the curve at certain frequencies. At each of these frequencies an ionospheric layer is penetrated. The three such discontinuities in the figure thus give the three critical frequencies for the three regions of maximum ionization E, F_1 and F_2.

The formation of these regions being due to the ionizing action of the solar ultraviolet radiation, it is obvious that the heights and the densities of ionization will vary greatly according to the hour of the day and the season of the year. Average values of the minimum virtual heights of the three regions in summer at sub-tropical latitudes during an epoch of maximum solar activity are 100, 210, 260 km. The corresponding number densities are 5×10^5, 10^6, $2{\cdot}5\times10^6$ electrons/cm.^{3}

* The terms 'Region' and 'Layer' are not infrequently used indiscriminately to denote a portion of the ionosphere having some characteristic feature such as a maximum of ionization density. The Wave Propagation Committee of the Institute of Radio Engineers, U.S.A., has, however, defined these terms as follows:

Region—A region of the ionosphere is a portion of the atmosphere in which there is a tendency for the formation of definite ionized layers.

Layer—A layer of the ionosphere is a regularly stratified distribution of ionization which is formed in a region of the ionosphere and is capable of reflecting radio waves back to earth.

It should be noted that it is not impossible for other stratifications to exist between these three regions which we are unable to detect by the ordinary experimental methods [56a]. It is obvious from Fig. 39 that the ionization of the various layers increases progressively from the lowermost one. If there is a stratum of maximum ionization, say, between F_2 and F_1 such that its maximum density is less than that of F_1, the stratum will not be able to produce any echo. Because, a wave which has penetrated the F_1-layer will also obviously penetrate this other layer. The experimental method described above, therefore, only indicates the existence of strata, the electron number density of any one of which is greater than that of any existing below.

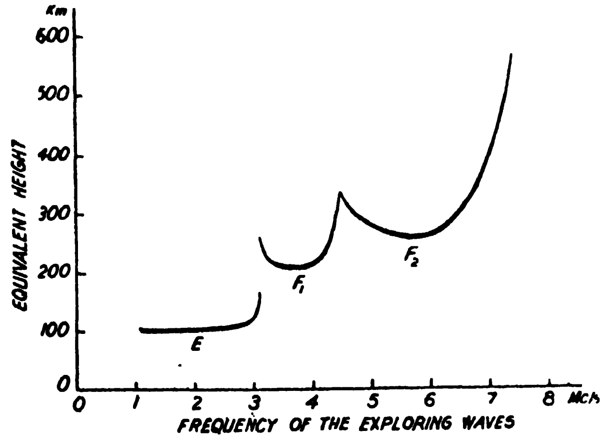


FIG. 39. Illustrating typical h'—f record indicating the stratification of the ionosphere.

Mention may be made in this connection of a probable G-layer situated above F_2 whose density is supposed to be ordinarily less than that of the F_2-layer. The G-layer is thus detectable only under exceptional circumstances when the density of F_2-layer may have fallen below that of this layer. The height of the layer is estimated to lie between 500 and 600 km. [57, 58]. It should, however, be mentioned that the occasional observations of such large values of equivalent height have also been interpreted as due to multiple reflection between the bottom of F_2-layer and the top of F_1-layer. The existence of G-layer is thus still problematic.

The 'pulse' technique reveals, besides the well-recognized stratifications E, F_1 and F_2, regions or 'patches' of intense ionization close to the height of the E region. On account of their irregular occurrence these are known as *sporadic* E (designated as E_s) [59, 60, 61]. The nature and origin of E_s are not yet fully understood and will be further discussed in Sec. 13(b).

The existence of Region D is not revealed by the ordinary pulse technique. This is because of the inability of this region to reflect medium and short waves at vertical incidence. Nevertheless, various effects connected with the variation of signal strength of waves reflected

from E and F regions (due to their passage through this region) as also variation of the strength of atmospherics from daylight to night hours indicate the existence of the absorbing D-layer [62, 63, 64].

The most casual radio listener notices an increase of atmospherics with the fall of darkness and simultaneously with it an enhancement of the signal strength of medium wave stations at large distances. The natural explanation of this phenomenon is that during hours of sunlight the comparatively dense atmosphere below Region E is ionized, where, on account of frequent collisions of electrons and ions with the neutral gas molecules, radio waves are strongly absorbed. When the solar radiation is withdrawn, the ionization decays quickly on account of rapid recombination, and the absorption of radio waves ceases.

The D-layer is also able to reflect back very-long waves (waves of length greater than 10 km.) incident on it. The return of the waves, however, takes place not by a process of gradual bending, as in Fig. 2, but by a process more akin to reflection from a boundary discontinuity. This is because for such long waves ionization density changes by a significant amount within a height comparable to one wavelength (see Sec. 3). The virtual height of reflection of these waves is found to lie between 70 and 80 km. [65]. It is, therefore, sometimes suggested that the D region ionization lies between 70 and 90 km. [66]. It should, however, be remembered that the observed height of reflection is the virtual height rather than true height. The true height of reflection may be perceptibly lower as has been calculated by Pfister [67]. The maximum electron number density of D region is estimated to be some few thousand electrons per cm.^{3} Reflections at vertical incidence from a virtual height of about 65 km. have also been reported by some workers [64, 68]. The ionized layer which produces such reflection is, however, to be distinguished from the normal D region (discussed above) and may be called sporadic D. The normal D causes absorption but is important in long wave propagation. The D-layer will be further discussed in Sec. 10.

(e) Group path and equivalent path—A theorem on equivalence of virtual heights of reflection for oblique and for vertical incidence

As mentioned earlier, the quantity measured in the group retardation method for finding out the heights of ionospheric layers is the time taken by the pulses to travel along the curved path TOR (Fig. 40), which may be called the *group path*. This time, as was pointed out, may be proved to be equal to the time which the wave group would take to travel along TAR, the whole of the path being considered to be in vacuum. The path TAR is therefore referred to as the equivalent path.

Consider the part of the group path MON lying within the ionospheric layer. The velocity v of the group varies from c at M, the point of entry

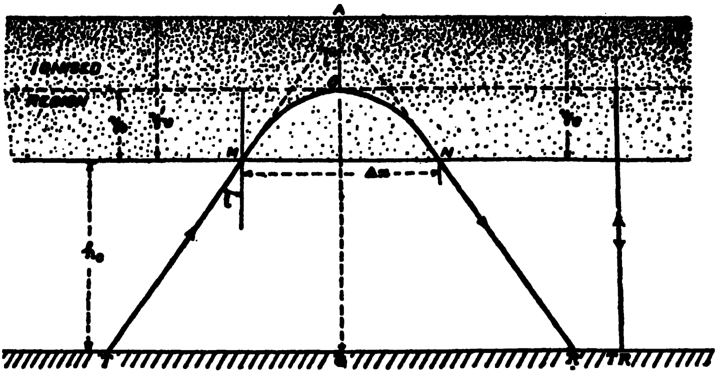


FIG. 40. Illustrating the equivalence of oblique incidence and vertical incidence equivalent heights.

into the layer, to zero at O the point of reflection. The total retardation Δt of the group is given by

$$\Delta t = \int_{MON} \frac{ds}{v}.$$

Or, since,
$$v = \frac{c^2}{u} = c\mu,$$

where u is the phase velocity and μ the refractive index,

$$\Delta t = \frac{1}{c}\int_{MON} \frac{ds}{\mu}.$$

If ψ is the angle made by the direction of propagation of the group at any point in its path with the vertical, μ_ψ the refractive index at this point and ds is any element of the path, then dx, the component of ds in the horizontal direction, is given by $ds = dx/\sin\psi$. Now, if i is the angle of incidence of the ray on the ionosphere then, by Snell's Law $\mu_\psi \sin\psi = \sin i$.

Hence,
$$\Delta t = \frac{1}{c}\int_{MON} \frac{ds}{\mu_\psi} = \frac{1}{c}\int_0^{\Delta x} \frac{dx}{\mu_\psi \sin\psi} = \frac{1}{c}\int_0^{\Delta x} \frac{dx}{\sin i}$$

$$= \frac{\Delta x}{c \sin i} = \frac{MAN}{c}, \qquad \text{.. .. ..} \qquad (53)$$

since, from geometry, $MAN = \Delta x/\sin i$. In other words, the measured retardation Δt for transmission of the group along the curved path MON is the same as the time taken by it to travel along the path MAN with the constant velocity c.

Another important theorem follows directly from the theorem of equivalence of path proved above.

Consider in the ionized region (horizontally stratified) a layer of electron number density N_e^0 and let a wave of frequency f incident normally on the region (from below) be reflected from this layer. Let another wave of frequency f' incident at an angle i on the region be also totally reflected from the *same* layer of electron number density N_e^0. Then, according to this theorem, *the equivalent heights reached by the two waves are also the same.* The proof of this theorem is as follows:

Let N_e^0 be the electron density in the region O, from which a wave of frequency f' incident obliquely on the ionosphere at an angle i, is reflected back (Fig. 40). Then we have

$$\sin^2 i = \mu_0{}^2 = 1 - \frac{N_e^0 e^2}{\pi m f'^2}$$

or,
$$f'^2 \cos^2 i = \frac{N_e^0 e^2}{\pi m} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (53a)$$

where μ_0 is the refractive index of the reflecting layer for frequency f'.

For reflection of a wave of frequency f incident normally from the same layer, we have

$$1 - \frac{N_e^0 e^2}{\pi m f^2} = 0$$

or,
$$f^2 = \frac{N_e^0 e^2}{\pi m} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (54)$$

From (53a) and (54), the two frequencies f and f' are related by the equation

$$f = f' \cos i. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (55)$$

f is called the *equivalent normal incidence frequency* corresponding to the oblique incidence frequency f'.

The refractive index μ_ψ of a region of electron density N_e for the wave frequency f' may thus be written in terms of the equivalent normal incidence frequency f as

$$\mu_\psi^2 = 1 - \frac{N_e e^2}{\pi m f'^2} = 1 - \frac{N_e e^2}{\pi m f^2} \cdot \cos^2 i.$$

From the relation $\mu_\psi \sin \psi = \sin i$, we also have

$$\mu_\psi^2 = \mu_\psi^2 \cos^2 \psi + \sin^2 i.$$

Equating the two expressions for μ_ψ we get

$$\mu_\psi^2 \cos^2 \psi = \cos^2 i \left(1 - \frac{N_e e^2}{\pi m f^2}\right) = \cos^2 i \,.\, \mu^2, \quad \ldots \quad \ldots \quad (56)$$

where μ is the refractive index of the same region for wave frequency f. Now we have from Fig. 40 and Eq. (53),

$$y_e' = \cos i \cdot \frac{MAN}{2}$$

$$= \frac{\cos i}{2} \int_{MON} \frac{ds}{\mu_\psi}.$$

If dy is the component of the path-element ds in the vertical direction, $dy = ds \cos\psi$ and therefore,

$$y_v' = \frac{\cos i}{2} \cdot 2 \int_0^{y_0} \frac{dy}{\mu_\psi \cos\psi},$$

y_0 and y_v' are the actual and equivalent heights of reflection reckoned from the lower boundary of the ionized region, for oblique incidence. From Eq. (56),

$$y_v' = \cos i \cdot \int_0^{y_0} \frac{dy}{\mu \cos i} = \int_0^{y_0} \frac{dy}{\mu}.$$

The integral on the right gives the equivalent height y_v of reflection of the wave at normal incidence, so that

$$y_v' = y_v.$$

Thus, we arrive at the important theorem [69] that *the equivalent height of reflection of a wave of frequency f incident vertically on the ionosphere is the same as that of a wave of frequency f′ incident obliquely, provided that the two waves are reflected from the same region of ionization.*

It should be remembered that the above theorem holds good only for a flat earth and a horizontal ionosphere. If the curvature of the earth is taken into account, then the ionospheric layer is also necessarily curved and a correction has to be applied.

(*f*) The Lorentz polarization term

In deriving Eq. (6) [Sec. 2(*a*)] it has been assumed that the accelerating force on an electron in the ionosphere due to a radio wave is only that due to its electric vector. This is in accordance with Sellmeyer theory of dispersion in which no account is taken of the force arising out of the electric polarization of the medium due to the applied electric force. Hartree [70], however, pointed out that in the case of the ionosphere the effect of polarization ought to be taken into account and that a term equal to $4\pi lP$ (where P is the polarization moment and $l = \frac{1}{3}$) should be added to the electric field of the wave. This is obviously in accordance with the Lorentz theory of polarization. The question whether Lorentz theory ($l = \frac{1}{3}$) or Sellmeyer theory ($l = 0$) is to be applied in ionospheric calculations is of considerable practical importance. For instance, it can be shown that the use of the one or the other may result in a variation of 50 per cent in the value of the electron density calculated from the critical penetration frequency.

The question whether the polarization term ought or ought not to be added has been closely examined by Darwin [19] and he gives strong theoretical reasons in favour of Sellmeyer theory ($l = 0$).

It is possible to test experimentally the validity of the one or the other theory by certain ionospheric measurements. Such measurements have

been made by various workers but the results obtained appear to be somewhat conflicting. There still remains some uncertainty as to the correct formula to be applied, though, on the whole, the conclusion is in favour of omitting the polarization term. We give below a short account of the experimental tests and of the results obtained.

One experimental method of deciding between the two theories depends upon the fact that for wave frequencies less than the gyromagnetic frequency and under certain conditions of propagation, the behaviour of the extraordinary wave predicted by the one theory is different from that predicted by the other theory [71]. If we consider vertical incidence radio exploration of the ionosphere at places where the magnetic dip is greater than 35°, and with wave frequencies for which vertical propagation is of the quasi-transverse type [Sec. 2(*d*)] then we have the following:

According to Sellmeyer theory	According to Lorentz theory
(i) Virtual height of reflection of the extraordinary wave will increase to infinity as the exploring wave frequency is increased to the gyromagnetic frequency.	(i) Virtual height of reflection of the extraordinary wave will increase to infinity as the exploring wave frequency is increased to the so-called Lorentz frequency, which is somewhat less than the gyromagnetic frequency.
(ii) Absorption suffered by the waves on frequencies near the gyromagnetic frequency will be very large.	(ii) Absorption of the extraordinary wave on frequencies near the Lorentz frequency will be less than that according to the Sellmeyer theory near the gyromagnetic frequency.

The ratio between the Lorentz frequency and the gyromagnetic frequency depends on the magnetic dip I of the place of observation and increases from 0·67 for $I=90°$ to unity for $I=35°$.

Now, Booker and Berkner [72] have analyzed a large number of $(P'—f)$ curves for Region F observed at night at Washington and have found that on frequencies far below the usual critical frequency, the extraordinary echo attains very large virtual heights increasing to infinity at a frequency of 1·38 Mc/s. Further, right up to this frequency the absorption is quite small as evidenced by the occurrence of multiple reflections. A typical record showing this phenomenon is reproduced in Fig. 41. Now, since the gyromagnetic frequency at Washington for Region F is 1·53 Mc/s and the value of the Lorentz frequency is 1·38 Mc/s, it appears that the observations agree with the predictions of Lorentz theory. Booker and Berkner maintain, therefore, that these observations are sufficient proof of the

validity of Lorentz theory in ionospheric calculations as against Sellmeyer theory.

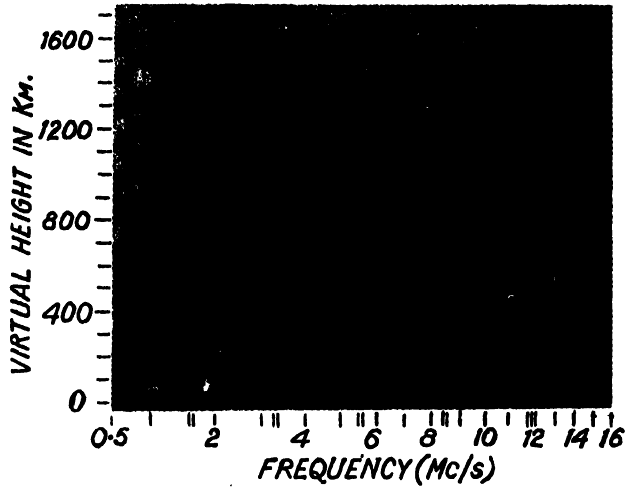


FIG. 41. Record of (P'—f) curve showing large virtual heights of Region F echoes (extraordinary) below Lorentz frequency (Booker and Berkner).

On the other hand, Martyn and Munro [73] made some interesting observations, which receive simple interpretation on the basis of Sellmeyer

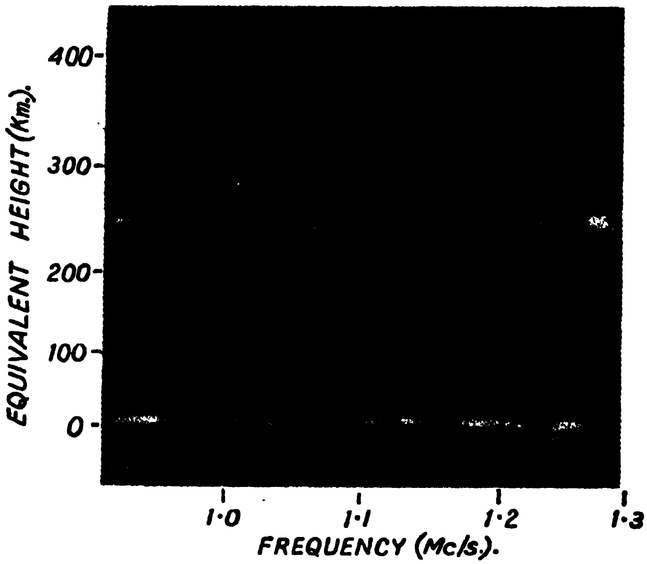


FIG. 42. Record of (P'—f) curve showing cessation of echoes on wave-frequency $f = f_H \cos\theta$ (Martyn and Munro).

theory. They frequently observed at night at Sydney a type of (P'—f) curve for Region F as shown in Fig. 42. The two uppermost traces in this

figure are interpreted by Martyn and Munro on the basis of Sellmeyer theory as follows:

(i) Both the traces are due to ordinary echoes; the one producing the lower trace is due to reflection according to the condition

$$p_0^2 = p^2 \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \quad (28)$$

and the other, producing the upper trace, is due to reflection according to the fourth condition of reflection [Sec. 2(c)],

$$p_0^2 = p^2 \frac{p^2 - p_H^2}{p^2 - p_L^2} \ldots \qquad \ldots \qquad \ldots \qquad \ldots \quad (30)$$

Eqs. (28) and (30) reduce to

$$N_{\max} = \frac{\pi m f^2}{e^2} \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots (28.1)$$

and

$$N_{\max} = \frac{\pi m f^2}{e^2} \cdot \frac{f_H^2 - f^2}{f_H^2 \cos^2\theta - f^2} \qquad \ldots \qquad \ldots \qquad \ldots (30.1)$$

where f is the exploring wave frequency $(=p/2\pi)$, f_H the gyromagnetic frequency $(=p_H/2\pi)$ and all other symbols have the same meaning as in Sec. 2. These equations, it should be noted, do not hold if the Lorentz polarization term is included [20] in the dispersion formula.

Since at Sydney the value of the gyromagnetic frequency for Region F is 1·45 Mc/s., the case of p less than p_H (y greater than 1) holds and it will be seen from Fig. 13 that the waves reflected according to Eqs. (28.1) and (30·1) are both ordinary waves.

(ii) It will be noticed in Fig. 42 that the uppermost trace continues up to about 1·27 Mc/s. and then disappears completely.

Eq. (30·1) shows that as the exploring wave frequency approaches $f_H \cos\theta$, the electron number density required to reflect the ordinary wave corresponding to the uppermost curve increases rapidly and finally becomes infinite when $f=f_H \cos\theta$. It is clear therefore that under this condition reflections will cease.

At Sydney the value of $f_H \cos\theta$ for Region F (230 km.) is 1·30 Mc/s. The disappearance of the uppermost trace just below 1·30 Mc/s. is therefore in striking agreement with Sellmeyer theory. Since, as has been pointed out above, the inclusion of the Lorentz polarization term does not give the condition of reflection represented by Eq. (30·1), i.e., does not predict cessation of reflection just below 1·30 Mc/s., Martyn and Munro maintain that the above observations are definite proof that Sellmeyer theory, and not Lorentz theory, applies to the case of the ionosphere.

Similar (P'—f) curves obtained by Appleton, Farmer and Ratcliffe [74] in England have, however, been interpreted in a different manner. These authors made polarization measurements [75] of the two waves and found that the upper curve is due to extraordinary reflection and the lower curve to ordinary reflection. They therefore pointed out that unless similar polarization measurements were made of Martyn and Munro's echoes, it is difficult to accept their interpretation that both the waves are ordinary.

Martyn and Munro [73] in subsequent investigations measured the polarization of the two echo components on about 500 occasions. They found that the echo corresponding to the uppermost trace had either left-handed elliptical or approximately linear polarization, but was never found to be circularly polarized. These results they maintain, uphold their interpretation and prove the validity of Sellmeyer theory.

Yet another method of testing the validity of the one or the other theory is provided by measurements of maximum usable frequency over a given transmission range (Sec. 9). The effect of including or omitting the Lorentz polarization term ($l=1/3$) in calculations on the transmission of radio waves through an ionospheric region was first discussed by Ratcliffe [76] and later by Smith [77]. Assuming the region to have parabolic distribution of ionization density with height and neglecting the effects of the collisional frequency of electrons and of the earth's magnetic field, Ratcliffe came to the conclusion that for transmission ranges round 500 km. the values of the maximum usable frequency for Region F_2 calculated according to Sellmeyer theory and according to Lorentz theory, differ only by about 2 per cent. For shorter and longer distances of transmission the predicted maximum usable frequencies on the two theories have no significant difference. Farmer, Childs and Cowie [78] made some simultaneous observations of vertical incidence (P'—f) curves and of the maximum usable frequency for a transmission distance of 464 km. using Region F_2 as the reflecting layer. Ratcliffe made use of these observations and showed [76] that omission or inclusion of the Lorentz polarization term meant a difference of only about 1 per cent. However, such small differences cannot be relied upon for discriminating between the theories, specially because the accuracy of these observations was about of the same order, 0·5 per cent.

Later analysis by Smith [77] shows, however, that for distances much larger than those used by Ratcliffe, the difference in the maximum usable frequencies calculated according to the two theories is much greater. For instance, it is found that for transmission by Region F_2, this difference amounts to about 12 per cent for a transmission distance of 1,000 km. and increases to a limiting value of 17 per cent for distances greater than 3,000 km. For transmission by Region E, this difference is nearly 8 per cent for a transmission distance of 500 km. rising to a limiting value of about 19 per cent for distances greater than 1,500 km. Smith's results have been applied to observed data on oblique incidence radio transmission and it has been found that the records show good agreement with Sellmeyer theory ($l=0$). More direct observations on maximum usable frequencies made by Beynon [79] also provide strong evidence in favour of the Sellmeyer theory. He has made accurate measurements of maximum usable frequency for transmission distances of 1,000 km. and 700 km. and compared the results with the values calculated from the two theories. It is found that for transmission over 700 km. the mean deviation of the observed values from the values obtained from Sellmeyer theory is only 1/10 of the

deviation from the values predicted by Lorentz theory. For transmission over 700 km. the divergence between the observed results and the values of maximum usable frequency calculated according to Sellmeyer theory is only 1/30 of what it would be if Lorentz theory is applied in the calculations. Sellmeyer theory is, therefore, in decidedly better agreement with observations on oblique incidence transmission than the Lorentz theory.

Thus the experimental evidence, on the whole, seems to be more in favour of omitting the Lorentz polarization term ($l=1/3$) from the dispersion formula.

It may be mentioned that in recent years Kelso [80] has made a detailed theoretical study of the effect of the Lorentz polarization term on the absorption of vertically incident wave in a deviating ionospheric region and has compared the results with those obtained from Sellmeyer's theory. According to the author, it is possible, with the help of the theoretical analysis made by him and with sufficient precision measurement, to determine whether the Lorentz term should be retained or not.

5. ABSORPTION IN AN IONIZED LAYER—REFLECTION COEFFICIENT

(a) Introduction

The study of the absorption of radiowave in its passage through an ionized layer is of great practical importance because all long distance radio communication is maintained *via* the ionosphere.

The absorption in the ionosphere may conveniently be classified as: (i) absorption in the deviating region, i.e. absorption suffered in the process of 'reflection' and (ii) absorption in the non-deviating region, i.e. absorption suffered in the part of the ray path where it is straight. Remembering that long distance propagation takes place by reflection from the *E*- or the *F*-layer, experimental measurements show (*vide infra*) that though there may be a small absorption loss in the process of reflection (particularly when the critical frequency is approached) by far the largest loss takes place in the underlying non-deviating *D*-layer, when this layer is present (as in day-time). Besides these ionospheric losses there is also, in the case of multi-hop transmission, loss at ground reflection. It is customary to lump together all these losses and speak of an 'overall absorption'. Corresponding to the coefficient of 'overall absorption' we also speak of the 'apparent coefficient of reflection'. The overall absorption suffered by radio waves in their journey to the level of reflection and back may be obtained from the apparent reflection coefficient ρ of the ionosphere.

(b) Theoretical considerations: Relative absorptions of the ordinary and the extraordinary rays

We will now deduce some relations defining the apparent reflection coefficient and overall absorption coefficient. Results of experimental studies will then follow.

If E_0 be the amplitude of the incident wave and E_1 that of the reflected wave, the apparent reflection coefficient ρ is given by

$$\rho = \frac{E_1}{E_0}. \qquad \dots \quad \dots \quad \dots \quad \dots \quad (57)$$

As the echoes return from the ionosphere by a process of refraction rather than that of reflection, the above ratio can be expressed as

$$\frac{E_1}{E_0} = \exp\left[-\int k\,dh\right] \qquad \dots \quad \dots \quad \dots \quad (58)$$

where k is the absorption coefficient of the medium through which the wave passes and dh an element of its path, the integration being carried out over the whole length of the path. So that

$$\rho = \exp\left[-\int k\,dh\right] = \exp(-\eta) \quad \dots \quad \dots \quad \dots \quad (59)$$

$$\text{or,} \qquad \eta = -\log \rho$$

where η is the total attenuation of the waves.

In order to evaluate ρ or $\int k\,dh$, we may consider separately the losses suffered in the two regions—the non-deviating and the deviating. In Sec. 2(d), formulae have already been derived from which the absorptions of the ordinary and the extraordinary waves for the two cases may be calculated. In order to apply them, however, it remains to be ascertained to which of the two types—quasi-longitudinal or quasi-transverse—the propagation in the deviating and in the non-deviating regions belongs. This will evidently depend upon the value of the magnetic dip at the place of observation. As a typical example, we discuss below the case for Calcutta where H=0·435 gauss and magnetic dip = 31° 45′, which is also generally applicable to temperate latitudes.

Remembering that the propagation is quasi-transverse or quasi-longitudinal according as [Sec. 2(d)],

$$\left|\frac{\gamma_T^4}{4\gamma_L^2}\right| > \text{ or } < \left|(1+\alpha+j\beta)^2\right|,$$

we note that for vertical propagation, the inclination of the direction of propagation to the magnetic field is θ=58° 15′, and, therefore,

$$\left|\frac{\gamma_T^4}{4\gamma_L^2}\right| = \left|\frac{\gamma^4 \sin^4\theta}{4\gamma^2\cos^2\theta}\right| = \left|\frac{\gamma^2}{4\times 0{\cdot}53}\right| = \left|\frac{\gamma^2}{2}\right| \text{ approximately.}$$

For the ordinary component, reflection condition is $1+\alpha = 0$, i.e., $p^2 = p_0^2$, and

$$\left|\frac{\gamma^2}{2}\right| = \frac{1}{2}\left(\frac{p_H}{p}\right)^2, \text{ because } \gamma = \frac{pp_H}{p_0^2}.$$

Hence, in the deviating region the propagation is quasi-transverse or quasi-longitudinal according as

$$\frac{1}{2}\left(\frac{p_H}{p}\right)^2 > \text{ or } < |\beta^2|,$$

i.e., according as

$$\frac{1}{2}p_H^2 > \text{ or } < \nu^2, \left(\text{since } \beta = \frac{p\nu}{p_0^2} = \frac{\nu}{p}\right).$$

Now, for Calcutta,

$$p_H \doteqdot 8\times10^6;\ \frac{1}{2}p_H^2 = 32\times10^{12};$$

and $\nu \doteqdot 10^5$ in Region E, and 10^3 in F. For the ordinary wave, therefore, $\frac{1}{2}p_H^2 > \nu^2$. The propagation is thus quasi-transverse in the deviating region whether it is located in Region E or Region F.

For the extraordinary component, reflection condition is $1+\alpha = \pm\gamma$. Hence propagation in the deviating region is quasi-transverse or quasi-longitudinal according as

$$\left|\frac{\gamma^2}{2}\right| > \text{ or } < \left|(\pm\gamma+j\beta)^2\right|,$$

or

$$\left|\frac{\gamma^2}{2}\right| > \text{ or } < \left|\gamma^2+\beta^2\right|.$$

The expression on the right-hand side is always greater than $|\gamma^2/2|$ and, therefore, the propagation of the extraordinary wave in the deviating region is always quasi-longitudinal.

The modes of propagation and the absorption coefficients of the ordinary and the extraordinary waves in both the deviating and the non-deviating regions have been discussed in detail by Booker [31] and are shown in the table below.

TABLE II

	Deviating region	Non-deviating region
	Quasi-transverse	Quasi-longitudinal
Ordinary	$k_o = -\frac{p}{2c}\frac{\beta}{\alpha}\left(\frac{1}{\mu}-\mu\right)$	$k_o = \frac{p}{2c}\cdot\frac{\beta}{[\alpha-\lvert\gamma_L\rvert]^2+\beta^2}$
	Quasi-longitudinal	Quasi-longitudinal
Extraordinary	$k_e = -\frac{p}{2c}\frac{\beta}{\alpha+\lvert\gamma_L\rvert}\left(\frac{1}{\mu}-\mu\right)$	$k_e = \frac{p}{2c}\frac{\beta}{[\alpha+\lvert\gamma_L\rvert]^2+\beta^2}$

The total absorption $\int k\,dh$ and the value of the reflection coefficient ρ, therefore, may now be obtained by choosing the appropriate value of k for the region (deviating or non-deviating) and for the wave component (ordinary or extraordinary) concerned.

Absorption of extraordinary and ordinary waves in temperate latitudes.—For computation of long distance radio propagation characteristics and also for the identification of the echoes obtained in the group-retardation

experiments [Sec. 4(c)], it is necessary to know the relative absorptions suffered by the ordinary and extraordinary components in their passage through the ionospheric regions. One is tempted to estimate this by comparing the values of the absorption coefficient k for the two rays in the deviating and the non-deviating regions. But this is not generally correct. For, the two waves may travel along altogether different paths and in order to estimate the relative overall absorption of the ordinary and extraordinary rays, one has to compare the values of the integrated absorption coefficient $\int k\,dh$ instead of simply the values of k.

Hence, as

$$\frac{k_e}{k_o}=\frac{(\alpha-|\gamma_L|)^2+\beta^2}{(\alpha+|\gamma_L|)^2+\beta^2}=\frac{(p+|p_L|)^2+\nu^2}{(p-|p_L|)^2+\nu^2} \quad \ldots \quad \ldots \quad (59a)$$

i.e. as $k_e > k_o$ all along the path the integrated absorption coefficient of the extraordinary is greater than that of the ordinary. The extraordinary component is therefore more absorbed than the ordinary one.

For the deviating region, as already pointed out, one has to compare the values of $\int k_e\,dh$ and $\int k_o\,dh$ rather than simply the values of k_e and k_o. An approximate value of $\int k\,dh$ may be obtained as follows if the quantity Y/X is regarded as constant [Sec. 2(d)].

$$\int k\,dh = -\left[\int_{\mu \doteqdot 1}^{\mu \doteqdot 0} dh/\mu - \int_{\mu \doteqdot 1}^{\mu \doteqdot 0} \mu\,dh\right]\cdot 2\cdot\frac{Y}{X}\cdot\frac{p}{2c}.$$

The integration extends from the region where μ is nearly equal to 1, to where $\mu = 0$. The expression is multiplied by 2 to take into account the forward and the return journeys of the wave in the ionosphere.

Now

$$2\int_{\mu \doteqdot 1}^{\mu \doteqdot 0} dh/\mu$$

is equal to P', the group path of the wave and also

$$2\int_{\mu \doteqdot 1}^{\mu \doteqdot 0} \mu\,dh$$

is equal to the total optical path. Therefore,

$$\int k\,dh = -[\text{group path}-\text{optical path}]\cdot\frac{Y}{X}\cdot p/2c$$

$$= -[P'-P]\cdot\frac{Y}{X}\cdot p/2c.$$

Or, since $P \ll P'$, (since μ is less than unity, lying between 1 and 0),

$$\int k\,dh = -P'\cdot\frac{Y}{X}\cdot p/2c.$$

Substituting for Y/X, β/α for the ordinary ray and $\beta/(\alpha+|\gamma_L|)$ for the extraordinary ray, we have

$$\int k_o\, dh = -P_o' \cdot \beta/\alpha \cdot p/2c \quad \text{for the ordinary,} \qquad \text{(60)}$$

and $$\int k_e\, dh = -P_e' \cdot \frac{\beta}{\alpha+|\gamma_L|} \cdot p/2c \quad \text{for the extraordinary.} \qquad \text{(61)}$$

Therefore,

$$\frac{\int k_e dh}{\int k_o dh} = \frac{P_e'}{P_o'} \cdot \frac{\alpha}{\alpha+|\gamma_L|} = \frac{P_e'}{P_o'} \cdot \frac{p}{p-|p_L|} \qquad \text{(61a)}$$

If P_e' and P_o' are of the same order then the extraordinary wave is more absorbed than the ordinary.

Thus, for vertical propagation under the experimental conditions as prevail at Calcutta, and generally also in temperate latitudes in the northern hemisphere, the extraordinary is more absorbed than the ordinary both in the non-deviating and in the deviating regions of the ionosphere.

(c) Experimental Measurement of Ionospheric Absorption

As already stated, the overall absorption $\eta(=-\log \rho)$ can be obtained by measuring the apparent coefficient of reflection ρ. This latter can be obtained either by the 'pulse' method (as in Slough, England) or by the 'continuous wave' method (as in the C.R.P.L., Washington) [81, 83, 86].

(i) *Pulse Method.*—The apparent reflection coefficient ρ for a given angle of incidence can be estimated by a comparison of the amplitude of the direct pulse with that of the reflected downcoming pulse as received near the ground. Let

F'—the amplitude of the once reflected downcoming pulse at the receiver, and

d_1—the actual distance travelled by the downcoming pulse in a single up and down journey.

Then

$$F' = \frac{B\rho}{d_1} \qquad \text{(62)}$$

where ρ is the apparent reflection coefficient of the ionized region and B is a constant defining the amplitude of the pulse sent upwards by the transmitter.

If G is the amplitude of the direct pulse at the receiver, we may write

$$G = \frac{A}{d_0} \qquad \text{(63)}$$

where A depends upon the amplitude of the direct pulse sent by the transmitter and its attenuation by the ground and d_0 is the distance travelled by the pulse over the direct path,

Combining Eqs. (62) and (63) we get,

$$\frac{F'}{G} = \frac{Bd_0\rho}{Ad_1} \qquad \text{(64)}$$

or
$$\rho = \frac{A}{Bd_0} \cdot d_1 \cdot \frac{F'}{G} = C \cdot d_1 \cdot \frac{F'}{G} \qquad \ldots \quad \ldots \quad (65)$$

where C is a constant called 'the transmission coefficient' involving the characteristics of the transmitting aerial, the ground attenuation and the distance between the transmitter and the receiver.

Again, let

F''—the amplitude of the twice reflected echo, and

ρ_g—the reflection coefficient of the ground.

Then we have
$$F'' = \frac{B\rho^2\rho_g}{2d_1}, \qquad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (66)$$

remembering that the pulse has travelled the distance d_1 twice. If, to a first approximation, ρ_g be assumed equal to unity, we may write
$$F'' = \frac{B\rho^2}{2d_1}. \qquad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (67)$$

From Eqs. (62) and (67),
$$\rho = 2F''/F'. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (68)$$

It is thus possible to measure the reflection coefficient by simply measuring the ratio of the intensities of the twice and singly reflected echoes. The only equipment necessary for the purpose is a stable pulse receiver, as is used in ionospheric height measurements.

However, multiple reflections are not always present and as such it is necessary in practice to have methods for determining ρ even when a single echo is present. This can be done in the following way: Eliminating ρ between Eqs. (65) and (68),
$$C = \frac{2F''}{(F')^2} \cdot \frac{G}{d_1}. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (69)$$

d_1, F' and F'' are measured when there is multiple reflection, and at the same time the amplitude of the direct pulse G at the receiver is measured. The transmission coefficient C is thus evaluated. Having once determined C for a particular frequency, ρ can be calculated for the same frequency by measuring the amplitudes of the direct ground pulse and the once reflected echo and applying Eq. (65). For d_1 the equivalent path is taken. This of course is not strictly correct but the error introduced is not large compared to the error involved in estimating F'.

An alternative method, which may be employed when only one echo is present is to provide the receiver with a calibrated attenuator at the input of the i.f. stage and a reference amplitude mark on the screen of the oscilloscope. The attenuator is calibrated by observing the amplitude of the successive echoes in a series of multiple reflection. The relation between the amplitude of the received signal and $-\log \rho$ is thus found which can be used even when one echo is present.

(ii) *The Continuous Wave Recording Method.*—In the CW recording method, as the name implies, continuous wave is used. This method of recording the field strength and deducing ionospheric reflection coefficient therefrom has been extensively employed in the U.S.A.

The CW transmitter is situated some miles away from the recorder and operates on half-wave horizontal antennas a quarter-wave high. The receiver also operates from half-wave horizontal dipole antennas, a quarter-wave high and broadside to the transmitting antennas. For these antennas, the maximum radiation and the greatest pickup factors are for vertical sky-wave radiation. Continuous wave automatic field strength recorders are connected to the output of the receiver. It is clear that the field intensity as recorded at any point is that due to the combination of the different modes of propagation of the ordinary and the extraordinary waves, all of which fluctuate randomly in intensity with respect to each other. The resultant median field intensity for such combination is equal to the square root of the sum of the squares of the median values of all the components. However, the field intensity for one of the component waves—say ordinary, once reflected—may be determined from the composite record from a knowledge of the differential absorptions of the ordinary and the extraordinary waves and by the use of the root-sum-square rule. The field intensity thus obtained may also be expressed in terms of the apparent reflection coefficient of the wave. In the U.S.A., continuous records of field intensity and the measurement of the reflection coefficient of the ordinary wave therefrom are being made systematically for a number of wave lengths in the medium and short wave bands [81].

Besides the vertical incidence measurements of reflection coefficient by methods described above, oblique incidence measurements of the same have also been made [81, 82, 84]. However, according to Martyn [85] the oblique incidence absorption coefficient can be simply obtained from vertical incidence data by multiplying the latter by sec i_0 where i_0 is the angle of incidence of the wave at the absorbing region. Appleton, Beynon and Piggott [86] are of opinion that this simple rule requires some modification.

Since, as already shown, $\log \rho_e/\log \rho_o$ is equal to the ratio k_e/k_o (or $\int k_e\,dh/\int k_o\,dh$), a measurement of the former ratio enables one, with the help of Eqs. (59*a*) and (61*a*) to ascertain whether the main absorption has taken place in the deviating or in the non-deviating region. Observations by various workers, as already indicated, show that during daytime, for reflections from *E* and *F* regions the absorption is chiefly in an underlying non-deviating region (*D* region). For night-time reflection from the *F* region, for wave frequencies well removed from the critical frequency there is little absorption either in the deviating or in the non-deviating region. As the critical frequency is approached, there is absorption and that in the deviating region [87, 88, 89]. These results are simply interpreted, because in the night-time the *D* region disappears and for frequencies close to the critical frequency, the wave penetrates deep into the deviating 'reflecting' region.

6. ESTIMATION OF THE HEIGHTS OF MAXIMUM DENSITY OF IONOSPHERIC LAYERS; ASSUMPTION OF PARABOLIC GRADIENT—THE 'THICKNESS' OF A LAYER; THE SCALE HEIGHT

The 'pulse' method of exploring the ionized regions yields, as we have already seen, the so-called equivalent height of reflection which is always greater than the true height. Near the penetration frequency, the equivalent height may become enormously great, not giving any indication of the true height of the level of maximum density. It is, however, possible from an analysis of the recorded $(P'-f)$ curves to make an approximate estimate of the true height of reflection as a function of the wave-frequency [90, 91]. In practice, it is found expeditious to start by assuming that the ionization gradient has some simple geometric form. Quite simple analysis of the P'-f records then yields not only the height of the maximum density but also other parameters of the ionized layer.

Of the various possible distributions, the one with a 'parabolic' gradient near the maximum may be expected to agree most closely with the actual distribution according to the Chapman theory of simple layer formation. That this is so has been shown by Rydbeck [92], at least for Region F_2.

Neglecting the effect of the earth's magnetic field we can thus write

$$N_e = N_{\max}\left(1 - \frac{y^2}{4H^2}\right) \qquad \ldots \quad \ldots \quad \ldots \quad (70)$$

where N_e is the electron number density at any height y measured from the level of the maximum $N_{\max}$.

It will be shown in Sec. 10(*b*) that the constant $H = kT/mg$ is the so-called scale-height of the atmosphere in this region. This simple relation holds good to within 5 per cent in the region between the levels H above and H below the level of $N_{\max}$

Referring the heights to the ground level (Fig. 43) let

h_M—be the height of maximum electron number density,

h'—the height at which electron number density is N_e,

h_0—the height at which the electron number density begins to increase from zero; this height may be taken approximately as lying at the level $2H$ below the height of maximum ionization, i.e., $h_M = h_0 + 2H$.

From the above, we may also define the 'semi-thickness' of the parabolic layer as $y_m = h_M - h_0 = 2H$.

Eq. (70) may now be written as

$$N_e = N_{\max}\left[1 - \frac{(h_M - h')^2}{y_m^2}\right], \qquad \ldots \quad \ldots \quad (71)$$

or, if the heights y are measured from the lower 'boundary' of the region, that is, from the level h_0,

$$N_e = N_{\max}\left[2\,\frac{y}{y_m} - \frac{y^2}{y_m^2}\right]. \quad \ldots \quad \ldots \quad \ldots \quad (72)$$

Since the refractive index of the ionized medium at the level of N_e for a wave frequency f is given by $\mu^2 = 1-N_e e^2/\pi m f^2$, we have, from Eq. (72)

$$1-\frac{f_c^2}{f^2}\left[\frac{2y}{y_m}-\frac{y^2}{y_m^2}\right],$$

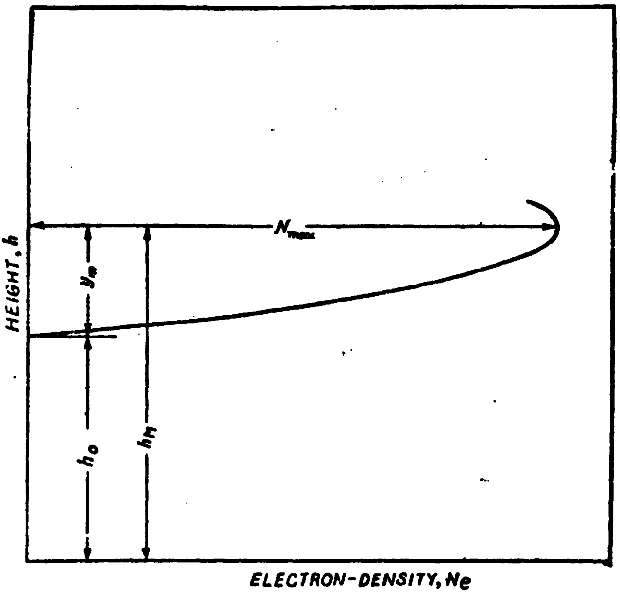


FIG. 43. Parabolic gradient of electron number density in the region of maximum ionization.

where f_c is the critical frequency of the layer so that N_{max} is equal to $\pi m f_c^2/e^2$.

Putting
$$a=\frac{2f_c^2}{f^2 y_m}, \text{ and } b=\frac{f_c^2}{f^2 y_m^2},$$

we get
$$\mu^2 = 1-ay+by^2. \qquad \ldots \quad \ldots \quad \ldots \quad (73)$$

This shows that μ is equal to zero for two values of y given by

$$y=\frac{a\pm\sqrt{a^2-4b}}{2b}. \qquad \ldots \quad \ldots \quad \ldots \quad (74)$$

Of these, the greater value corresponding to the positive sign before the radical may be neglected, for the wave cannot in general reach this height being reflected back from the lower height given by

$$y=\frac{a-\sqrt{a^2-4b}}{2b}.$$

The equivalent height h' reached by the wave is thus given by

$$h'=\int_{y=0}^{y=y}\frac{dy}{\mu}=\frac{1}{2}\log_e\frac{a+2\sqrt{b}}{a-2\sqrt{b}}.$$

$$=h_0+y_m\frac{f/f_c}{2}\log_e\frac{1+f/f_c}{1-f/f_c}\ldots \qquad \ldots \quad (75)$$

The above relation, in conjunction with the ionospheric data as obtained from $(P'—f)$ curves enables one to compute the characteristics of the layer, namely, the height of the level of maximum ionization and the 'semi-thickness' y_m of the layer. For this we may proceed as follows after Booker and Seaton [93].

Let us define a function

$$\phi\left(\frac{f}{f_c}\right)=\frac{f/f_c}{2}\log_e\frac{1+f/f_c}{1-f/f_c}-1$$

so that Eq. (75) can be written as,

$$h'=h_M+y_m\phi\left(\frac{f}{f_c}\right). \quad \ldots \quad \ldots \quad \ldots \quad (76)$$

In order to determine the two unknown quantities h_M and y_m, we require a pair of simultaneous equations derived from Eq. (76) with two different values of h' and $\phi(f/f_c)$. These are obtained by choosing two suitable points on the $(P'—f)$ curve, noting the values of h' and f/f_c for these two points and calculating $\phi(f/f_c)$ from the latter. For convenience of calculation a table such as given below may be prepared giving values of $\phi(f/f_c)$ for a set of suitable values of f/f_c.

TABLE III

f/f_c	0	0·648	0·725	0·757	0·834	0·887	0·901	0·925	0·969
$\phi(f/f_c)$	-1	$-\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{1}{2}$	1

An inspection of the table shows that for $(f/f_c) = 0·834$, $\phi(f/f_c)$ is zero. This provides a simple method of determining the level of maximum ionization as Eq. (76) shows that in this case $h' = h_M$. The equivalent height h' reached by an exploring wave of frequency f equal to $0·834f_c$ is thus equal to h_M, the height of the level of maximum ionization. To obtain the characteristics of an ionospheric region it is thus only necessary, (i) to measure h' for a frequency equal to 0·834 times the critical frequency of the region, and (ii) to note the value of h' for the portion of the $(P'—f)$ curve, where it is sensibly horizontal. The former gives the level of the maximum electron density and the latter the height of the 'bottom' of the layer h_0. The semi-thickness of the layer is obviously equal to h_M-h_0.

Fig. 44 depicts the variation of virtual height with exploring wave frequency as obtained from Eq. (75), i.e., the $(P'—f)$ curve for a region of parabolic distribution of electron density.

Booker and Seaton have extended their analysis to records made during daytime when all the three regions E, F_1 and F_2 are present. The ionosphere is then represented by a triplet of parabolas shown in Fig. 45. The corresponding $(P'—f)$ curves are shown in Fig. 33. (This distribution of

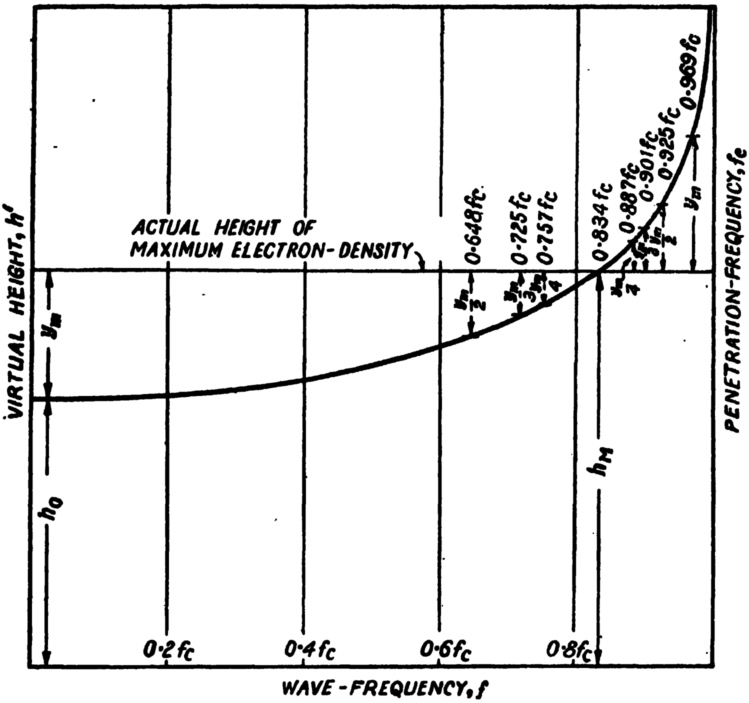


Fig. 44. Relation between virtual height and wave-frequency for reflection from a region of parabolic distribution. (After Booker and Seaton.)

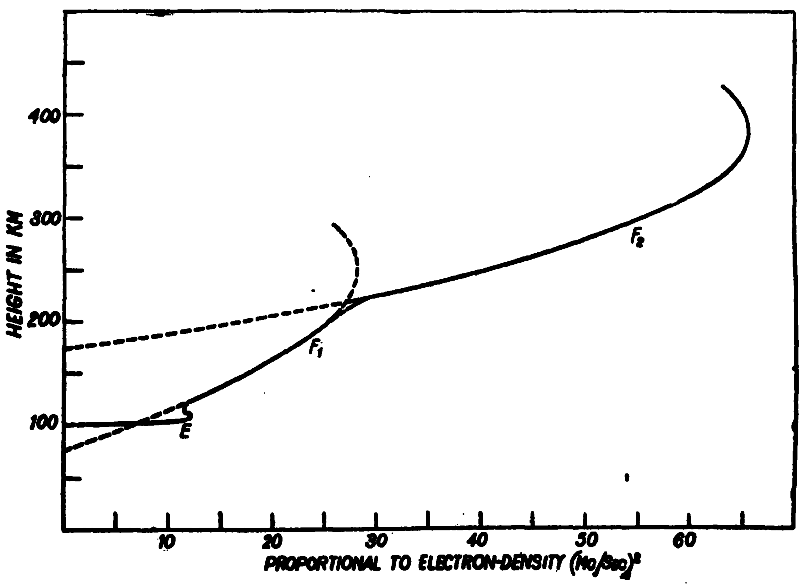


Fig. 45. Distribution of electron number density in the E, F_1 and F_2 regions derived from the (P'—f) curves in Fig. 33 on the assumption of parabolic gradient. (After Booker and Seaton.)

ionization may be compared with the distribution obtained theoretically by Bhar as depicted in Fig. 66.)

An alternative method of utilizing (P'—f) curves for determining the layer characteristics, on the assumption of a parabolic gradient, has been suggested by Appleton and Beynon [94].

We note from Eq. (75) that the relation between h' and

$$\frac{1}{2}\frac{f}{f_c}\log_e\frac{f_c+f}{f_c-f}$$

is a linear one and that the slope of the straight line obtained by plotting these two quantities gives y_m (Fig. 46). The point where the straight line cuts the h' axis thus gives the value of h_0.

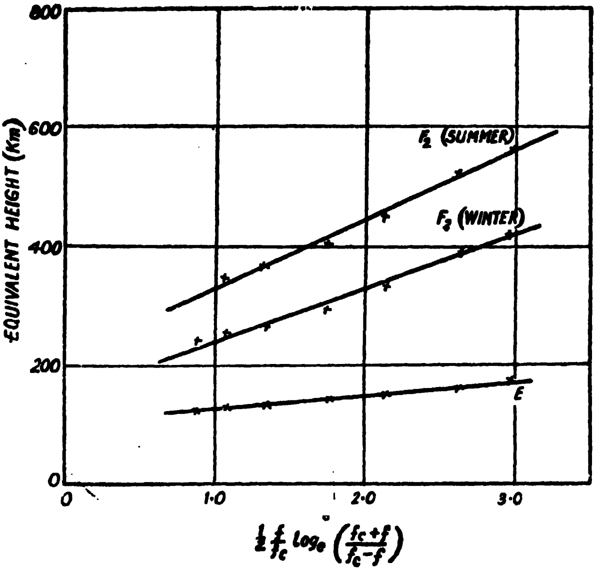


FIG. 46. Appleton and Beynon's method of determining h_0 (height of the bottom of the layer) from the linear relation between h' and $\phi(f/f_c)$.

It should be emphasized that all the above conclusions are based on the assumption of parabolic distribution of electron density and since, as already pointed out, the parabolic law holds only over a very limited range on either side of the maximum electron density, it follows that for accurate determination of h_0 and y_m, one should utilize only the portion of the (P'—f) curve near the critical penetration frequency.

According to Booker and Seaton [93] for the region below the maximum, the parabolic law holds down to the level at which the electron density is about 20 per cent of the maximum. This means that for the above calculations, the (P'—f) curves can be utilized down to a frequency equal to 0·447 times the critical penetration frequency.

Appleton and Beynon [94] are, however, of opinion that one should not use the (P'—f) curves for frequencies below 0·9 f_c.

In the determination of h_0 for an ionized layer one has to allow for retardation by underlying layers, if any. For instance this correction for Region F for an underlying E layer of semi-thickness y_E, should be

$$-\frac{2}{3}y_E\left(\frac{f_E^o}{f_F^o}\right)^2,$$

where f_E^o and f_F^o are the ordinary ray critical frequencies for Regions E and F respectively.

Since $y_m(=2H=2kT/mg)$ involves the temperature and mean molecular weight of the constituent gases in the ionospheric region concerned, determination of y_m is of considerable physical importance. Apart from this, a knowledge of the layer parameters h_0 and y_m is of great value in determining theoretically the characteristics of long distance radio communication such as, range, maximum usable frequency. These will be discussed in Sec. 9.

As already indicated the above results hold good only for a limited portion of the Chapman region near its maximum ionization. For the complete Chapman region, corrections to the layer parameters derived by Pierce [95] and given below may be used.

$$H = 0{\cdot}60\, y_{mp}$$

$$h_{MC} = h_{Mp} - 0{\cdot}14\, H$$

$$y_{mC} = 1{\cdot}10\, y_{mp}.$$

where the subscripts C and p refer to the Chapman and to the parabolic distribution respectively. It is to be noticed that the corrections have *reduced* the height of the maximum density and *increased* the value of the semi-thickness.

7. COLLISIONAL FREQUENCY OF ELECTRONS WITH ATOMS AND MOLECULES IN THE IONOSPHERIC REGIONS

(a) Introduction

We have already mentioned that collisions of electrons and ions with the gas particles cause absorption of radio waves passing through the ionosphere. The process of absorption is as follows. The advancing wave sets the electrons and ions in vibration and in doing so necessarily spends energy. In the absence of collisions (i.e., if the average interval between successive collisions is large compared to the period of the wave) this energy is restored to the advancing wave by the vibrating electrons sending out coherent radiation. If, however, there be frequent collisions within a complete period of vibration of the electron then the absorbed energy is not so restored. The radiation from the electrons due to sudden and random changes of the momenta is of a scattered nature and is lost so far as the advancing wave is concerned. There is then absorption of energy from the passing radio wave.

The amount of absorption thus depends upon the frequency of collisions and this in its turn, upon the cross-sections of the colliding particles. The meaning of the term cross-section, as applied to atoms and molecules having a very open structure of electric charges, should be clearly understood. Evidently, a structure like this cannot have a well-defined cross-section in the sense in which the term is used in the gas kinetic theory. In fact, the value of the cross-section depends upon the velocity and the nature of the reaction taking place between the colliding particles. The cross-section for producing ionization, for instance, is different from that for excitation or from that for elastic collision. The respective cross-sections for the above processes can be determined by quantum mechanical methods. For elastic collision in particular, with which we are concerned here, it has been found, both theoretically and experimentally, that the cross-section is not a constant quantity, but is dependent on the velocity of the colliding electron. In order to calculate the collisional frequency, it is, therefore, necessary to know how the cross-sections of the atmospheric constituents vary with the energy of the colliding electrons. It may be mentioned that experimental and theoretical data are available for the cross-sections of O_2 and N_2, [96, 98]. For O, theoretical calculations have been made [97]. Such results are, however, not available, for the corresponding positive ions.

(*b*) Theoretical Considerations

The collisional frequency ν obviously depends upon the mean thermic velocity ($\bar{v}$) and the mean free path of the electron (l_e) in the gas. We have

$$\nu = \frac{\bar{v}}{l_e} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (77)$$

The expression for $\bar{v}$ may be taken as

$$\bar{v} = \sqrt{8kT/\pi m} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (78)$$

Also, l_e may be obtained from the relation

$$l_e = \frac{1}{n\pi\sigma^2} = \frac{1}{nA} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (79)$$

where n is the number of atmospheric particles per cm.^{3} and $\pi\sigma^2$ is the particle cross-section A. The average value of the latter, as obtained from viscocity experiments in air is of the order 7×10^{-16} cm.^{2} l_e has also been obtained experimentally. Thus, according to Townsend and Tizard [100] and Elias [101], l_e corresponding to a temperature 0°C. is approximately given by $l_e = 3{\cdot}2 \times 10^{-2} \times 1/P$ cm., where P is the pressure in mm. of Hg. More recent measurements [99] give the value of l_e corresponding to a temperature 15°C. as $5{\cdot}4 \times 10^{-2} \times 1/P$.

Knowing the temperature and pressure in the region concerned, we can therefore calculate $\bar{v}$ and also l_e and hence estimate the collisional frequency. Thus calculated the collision frequencies of electrons in the E and F_2 regions are found to be of the order $1{\cdot}4 \times 10^5$/sec. and 70/sec. respectively. Of these two values the value for region E agrees well with that determined

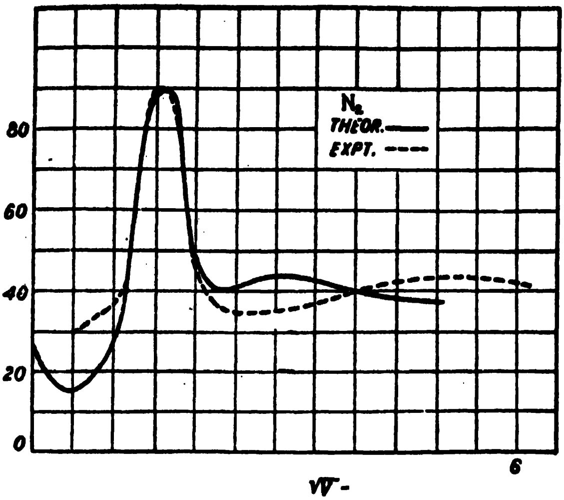


FIG. 47. Variation of the collisional cross-section of N_2 for elastic collision with electrons of different energies. (After Fisk.)

from radio measurements (*vide infra*). For Region F_2, however, the value is widely different. This discrepancy is to be traced to the fact that the gas kinetic cross-section (as assumed in the above calculation) does not represent the actual cross-section involved in the collision. As already mentioned the cross-section depends upon the energy of the colliding electron. When this is taken into account the discrepancy disappears. In what follows we shall discuss this aspect of the problem of collisional frequency in Regions E and F.

Figs. 47, 48, and 49 depict the variations of the cross-sections of N_2, O_2 and O respectively for elastic collision with electrons of different energies. For N_2 and O_2 both theoretical and experimental curves are given. For O only the theoretical curve as available is given. As already mentioned l_e

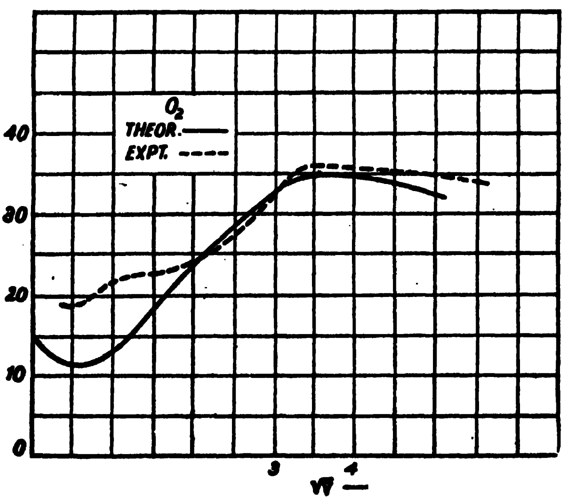


FIG. 48. Variation of the collisional cross-section of O_2 for elastic collision with electrons of different energies. (After Fisk.)

(which is a function of the cross-section) has been determined experimentally for air, that is mixture of N_2 and O_2.

For calculating ν in Region *E* we have the following data:

Region E (100 *km. level*)

Number density of N_2	$2{\cdot}8 \times 10^{13}/\text{cm.}^3$
Temperature	300°K.
Cross-section of N_2 (from Fig. 47)	20 atomic units.
Collision frequency	$1{\cdot}8 \times 10^6/\text{sec.}$

In the above calculation we have neglected collisions with O_2. This is because while the cross-section of O_2 is approximately the same as that of N_2 its number density is in all probability small compared to that of N_2. The number density of O is also assumed to be small. If instead of the cross-section from Fig. 49 we use the value of the mean free path as obtained from experimental results of Huxley and Zaazou [99, 99*a*] we get the value of ν to be 7×10^5/sec. for the bottom of Region *E* (height 88 km.). These values are in substantial agreement with those obtained from cross-modulation experiments (*vide infra*). However, the agreement with the value as calculated from the consideration of the gas kinetic cross-section is accidental rather than real, because, the gas kinetic cross-section happens to be of the same order as that of the actual cross-section for collisions with slow electrons. Generally the two values are different. This, as shown below, is for the case of atomic oxygen.

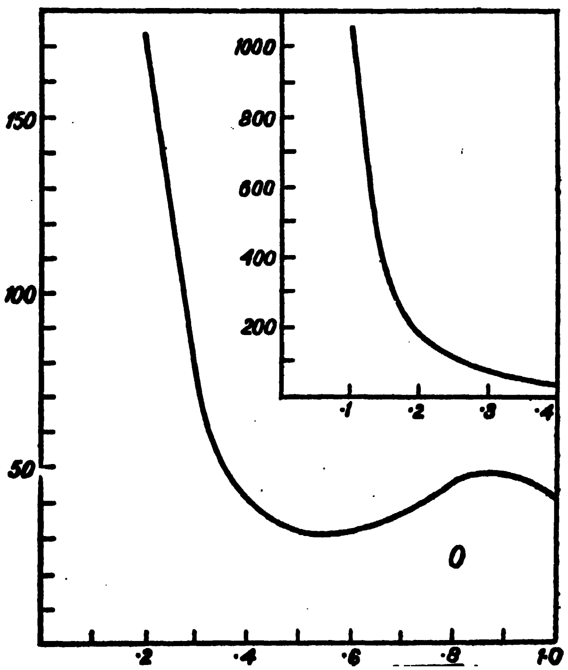


FIG. 49. Variation of the collisional cross-section of O for elastic collision with electrons of different energies.

Region F (250 *km. level*).

Number density of O-atoms	..	$2\cdot5\times10^{9}$/cm.^{3}
Temperature	..	1000°K.
Cross-section of O-atom	..	1060 atomic units.
Collisional frequency	..	$1\cdot5\times10^{3}$/sec.

In calculating the collisional frequency we neglect collisions with N_2. This is because the collisional cross-section of N_2 at 1000°K. is much less than that of O and its number density in the region under consideration is comparable to that of the O atoms. The value of ν thus obtained agrees well with the value 2×10^{3} per second obtained from radio measurements. The discrepancy, when the calculation is made with the gas kinetic cross-section, is now seen to be due to the fact that the effective 'elastic' cross-section calculated quantum mechanically is about 53 times the gas kinetic cross-section.

For lack of data it is not possible to calculate the frequency of collision between electrons and the positive ions N_2^+, O_2^+ and O^+. Since there are reasons to believe that in the upper atmosphere they exist in much smaller proportion than the neutral molecules and atoms, their contribution to the collisional frequency may not be important.

(*c*) Experimental determination of collisional frequency

(i) *From the phenomenon of interaction of radio waves* (*Luxembourg Effect*).—*Region E*: The phenomenon of cross-modulation in the ionosphere as a result of interaction of radio waves, offers a very interesting method of estimating the collisional frequency of electrons in Region E. This will be discussed in Sec. 15.

(ii) *From measurements of reflection coefficient and* (P'—f) *curves.*—*Region F*: An estimate of the collisional frequency in Region F can be made from observations on its reflection coefficient for wave frequencies not far removed from the critical frequency. The absorption in this case, as already mentioned [Sec. 5(*b*)], occurs in the deviating portion of Region F and the method of measurement consists in measuring simultaneously the equivalent path and the relative intensity of the reflected wave [102, 103, 62].

If we make the reasonable assumption that in Region F, $\nu^2 \ll p^2$ and that $(ck/p)^2 \ll \mu^2$, then we have [Sec. 5(*b*)]

$$k=-\frac{p}{2c}\cdot\frac{Y}{X}\left(\frac{1}{\mu}-\mu\right). \qquad (80)$$

For the ordinary wave, in the deviating region ($\mu \ll 1$), the propagation is of the quasi-transverse type and therefore

$$\left.\begin{aligned}X&=\alpha\\ Y&=\beta\end{aligned}\right\}. \qquad (81)$$

Substituting these values in Eq. (80) and remembering that $\alpha = -p^2/p_0^2$ and $\beta = pv/p_0^2$, we get,

$$k_o = \frac{\nu}{2c}\left(\frac{1}{\mu} - \mu\right). \qquad \ldots \quad \ldots \quad \ldots \quad (82)$$

(In Region F, $\nu \doteq 10^3$ per sec. and p is greater than $2\pi \times 3 \times 10^6$, so that the assumption $\nu^2 \ll p^2$ is justified. It can be easily seen that the condition $(ck/p)^2 \ll \mu^2 \ll 1$ is satisfied for values of μ as low as 10^{-2} and the above expression is valid up to the corresponding level.)

It can be seen from Eqs. (58) and (82) that

$$\log \rho_o = -\int k_o\, dh = -\frac{\nu}{2c}(P_o' - P_o) \qquad \ldots \quad \ldots \quad (83)$$

where, as before, P_o' is the group path and P_o the optical path of the ordinary ray. In observations as are being described here, the echo is much group retarded, i.e., $P_o' \gg P_o$. For the case of the ordinary wave, therefore

$$\log \rho_o = -\frac{\nu}{2c} P_o' \qquad \ldots \quad \ldots \quad \ldots \quad (84)$$

A linear relation thus holds between $\log \rho_o$ and P_o'. Hence, if simultaneous observations are made of ρ_o and P_o', ν may be determined from the slope of the curve $(\log \rho_o - P_o')$.

It is also possible [104] to estimate ν from the observed $(P' - f)$ curves even when P_o cannot be neglected in comparison to P_o'.

Let ρ_1, P_1, P_1' and ρ_2, P_2, P_2' denote the values of these quantities for the ordinary wave on two different frequencies f_1 and f_2 respectively; we can then write from Eq. (83)

$$f_1 \log \rho_1 = -\frac{\nu}{2c} f_1(P_1' - P_1) \quad \ldots \quad \ldots \quad \ldots \quad (85)$$

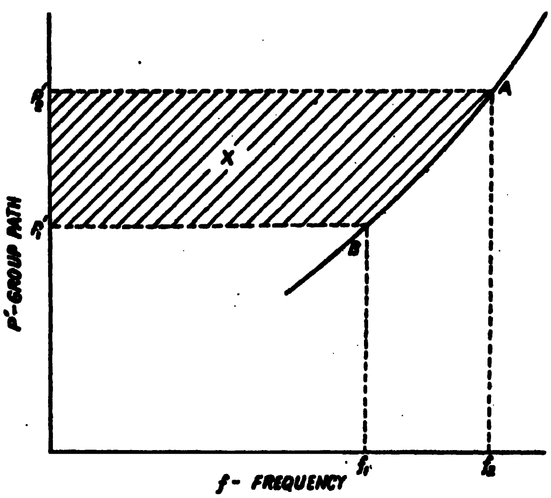


FIG. 50.

and

$$f_2 \log \rho_2 = -\frac{\nu}{2c} f_2(P_2'-P_2). \qquad \ldots \quad \ldots \quad (86)$$

Subtracting

$$f_1 \log \rho_1 - f_2 \log \rho_2 = \frac{\nu}{2c}[(P_2'f_2-P_1'f_1)-(P_2f_2-P_1f_1)] \quad \ldots \quad (87)$$

In Fig. 50 AB represents a portion of the observed $(P'-f)$ curve and the group paths P_1', P_2' correspond to the two wave frequencies f_1 and f_2. Then

$P_2f_2-P_1f_1$, which can be shown to be equal to $\int_{f_1}^{f_2} P'df$,

is represented by the area ABf_1f_2 and the quantity $P_2'f_2-P_1'f_1$ is given by the area $P_1'Bf_1f_2AP_2'$. Thus from Eq. (87)

$$f_1 \log \rho_1 - f_2 \log \rho_2 = \frac{\nu}{2c}[\text{area } P_1'Bf_1f_2AP_2' - \text{area } ABf_1f_2]$$

$$= \frac{\nu}{2c}[\text{shaded area } X];$$

so that

$$\nu = 2c \cdot \frac{f_1 \log \rho_1 - f_2 \log \rho_2}{\text{area } X}. \quad \ldots \quad \ldots \quad \ldots \quad (88)$$

Thus, if we determine the reflection coefficients of the ordinary ray on two frequencies, we can estimate the value of ν from the $(P'-f)$ curve with the help of the above relation.

Farmer and Ratcliffe [82] have estimated ν for Region F by this method and have found its value to be of the order $1{\cdot}2\times10^3$/sec.

Region E: For Region E, the changes in ρ and in group path and optical path are difficult to determine near the penetration frequency. Nevertheless, Briggs [83] with an improved type of ionospheric apparatus, has been able to estimate ν near the height 130 km. and has obtained the value 2×10^4/sec.

8. MEASUREMENTS OF POLARIZATION AND DIRECTION OF ARRIVAL OF DOWN-COMING WAVES

Various types of apparatus have been devised for studies of both polarization and direction of arrival of down-coming waves [105, 106, 107]. We will describe here the principle of a simple device due to Eckersley and Farmer which enables one to determine the polarization, and, at the same time, with a slight modification, the direction of arrival of the down-coming waves [108]. It can be used for the study of 'pulsed' echoes and embodies certain improvements over the polarization apparatus which had been devised earlier.

Two pairs of loop aerials—one north-south (N–S) and the other east-west (E–W)—arranged as shown in Fig. 51, are used in conjunction with

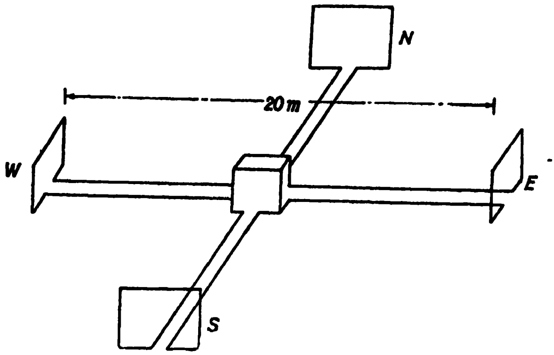


FIG. 51. Arrangement of loop aerials for measurements of polarization and direction of arrival of down-coming waves. (After Eckersley and Farmer).

two similar matched amplifiers (Fig. 53).

For polarization measurement, the N–S pair are coupled in parallel to one of the amplifiers and the E–W pair to the other. The outputs of the amplifiers are connected in the usual manner to the two pairs of deflecting plates of the C.R.O. If the incident wave is in the vertical direction then it is easy to see that the two pairs of aerials will behave as a single pair of crossed loops. If the two amplifiers are properly aligned, (i.e., if equal phase change is introduced between the input and output, and the amplification in one is exactly the same as that in the other), the spot on the C.R.O. screen will depict the actual polarization ellipse. In the case in which the waves arrive obliquely, there is a phase difference between the two frames in parallel. Account is taken of this in calculating the relative amplitude of the two fields. It is, however, to be noted that the phase relation of the two components of the polarization ellipse is not affected in any way even for oblique incidence.

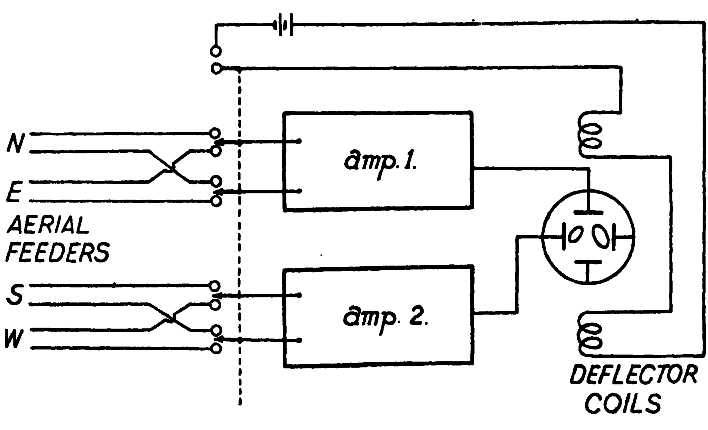


FIG. 52. Matched amplifiers and cathode-ray oscillograph used with the aerial system of Fig. 51.

For measurement of the direction of arrival we note that a down-coming wave, unless it is in the vertical direction, will, in general, arrive at the four loops at different instants and thus with different phases. For the special case when the plane of incidence is N–S (or E–W) the phases at the E–W (or the N–S) loops will be the same. The direction of arrival of the down-coming wave is thus determined if the phase relations between the N–S and the E–W frames are separately known. For this determination the N–S and the E–W aerials are alternately connected (via the two amplifiers), in quick succession, to the two pairs of deflecting plates of the C.R.O. In synchronisation with these alternations the cathode ray beam is deflected sideways by a pair of magnetic coils suitably placed. Two separate ellipses are thus produced. One of these indicates the phase relation between the N–S frames and the other that between the E–W frames. From these data the direction of arrival of the down-coming wave—both its elevation and azimuth—is completely determined.

The most important part of the experiment is the careful lining-up of the two amplifiers before commencement of the observations. This is done by picking up radiations from suitably placed aerials energised by local oscillators. Each of the frame aerials is also carefully aligned relative to its feeder, so that spurious pick-ups are not transferred to the frame. For fuller details the original paper should be consulted.

As mentioned earlier, the apparatus is suitable for the study of 'pulses'. Further, the apparatus can select a small element of any particular echo for study. This is achieved by means of an 'echo-selecting' device which keeps the C.R.O. de-sensitized for most of the time except only for very small periods of 15 μ sec. By means of suitable phasing circuits, the small active periods are made to coincide with those of the arrival of the particular echo of the pulse pattern which it is desired to study.

The authors have used the apparatus successfully for the study of various types of ionospheric echoes. In particular, they have studied the scatter echoes on high power transmissions (see Sec. 13(*c*)). These echoes are very short-lived and their characteristics change so rapidly that they cannot be followed by the usual manual methods.

9. APPLICATION OF IONOSPHERIC DATA TO RADIO TRANSMISSION

(*a*) Introduction

Records of routine observations made at the ionospheric observatories supply the following data for normal incidence of the exploring waves on the ionized layers: (i) the *equivalent* heights of the layers for all frequencies on which reflections are obtained, and (ii) the *critical frequencies of* the various ionized regions [*Vide infra*, sub-sec. (*d*)]. From these data for normal incidence, the following characteristics for oblique incidence for a given frequency of transmission may be obtained [94, 109, 110, 111]: (i) the maximum usable frequency (MUF) $f'_{\max}$ for transmission over a given distance D at any

time, and (ii) the skip distance D_{skip} for a given frequency. These characteristics are of extreme importance in the determination of the most suitable wave-lengths for long distance communication over a given path.

In England the Radio Research Board publishes the maximum usable frequency data as curves giving the relation between the maximum usable frequency factor $f'_{\max}/f_c$ (where f_c is the vertical incidence critical frequency) and the distance of transmission D for appropriate ranges of the parameters y_m and h_0 [94, 109]. (y_m is the semi-thickness of the ionized layer assumed to be of parabolic density distribution and h_0 is the height of the 'bottom' of the layer from the surface of the earth (Sec. 6).)

In Washington, at the National Bureau of Standards, the maximum usable frequency and the skip distance for each frequency are represented graphically for different hours of the day. These are determined by a graphical method due to Newbern Smith [110].

In what follows we will first describe the theory of drawing the curves after Radio Research Board, England. The graphical method of Smith will be briefly outlined later.

(*b*) Radio Research Board, England (Appleton and Beynon's method)

In the theoretical determination of the maximum usable frequencies it is convenient to consider separately two cases: (i) Distances within 500 km., for which the earth may be considered flat, and (ii) Distances greater than 500 km., for which the earth's curvature has to be taken into account. For each case two types of ionization gradient has to be considered. First, sharp ionization gradient (thin layer) and secondly low ionization gradient (thick layer). The first is applicable to reflection from E, sporadic E and F_1 layer. In this ideal case the ray does not penetrate into the layer. The second case holds for reflection from F_2 layer when a considerable part of the ray path lies within the layer (case of thick layer).

(i) FLAT EARTH—*Thin layer.* For the case of flat earth and thin layer the maximum usable frequency and the skip distance can easily be calculated geometrically. This case is practically realized when the distance of transmission is less than 500 km. and reflection takes place from a sporadic E region, say.

Neglecting collision and the effect of the earth's magnetic field, the maximum frequency f' which can be reflected from the layer at an angle of incidence i_0 is given by

$$\mu^2 = \sin^2 i_0 = 1 - \frac{N_0 e^2}{\pi m f'^2} = 1 - \frac{f_c^2}{f'^2}, \qquad \ldots \qquad \ldots \qquad (89)$$

where f_c—the frequency for which the refractive index of the medium becomes zero at the top of the trajectory, i.e. the critical frequency of the layer at vertical incidence,

N_0—maximum electron number density of the layer.

From Fig. 53 we have

$$\tan i_0 = \frac{D}{2h_0},$$

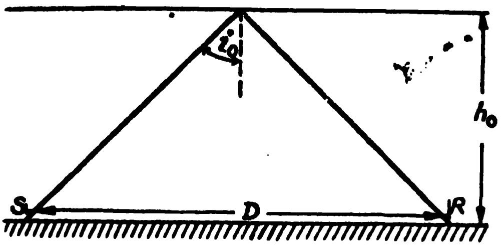


FIG. 53. Transmission path for the case of reflection from a thin ionized layer (flat earth).

where D is the distance of transmission and h_0 the height of the layer. Also, from Eq. (89)

$$\tan^2 i_0 = \frac{1-\dfrac{f_c^2}{f'^2}}{\dfrac{f_c^2}{f'^2}} = \frac{f'^2}{f_c^2} - 1.$$

Therefore

$$\frac{f'^2}{f_c^2} = \frac{D^2+4h_0^2}{4h_0^2}. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (90)$$

Eq. (90) gives the maximum frequency which may be used for a given distance or alternatively, the minimum distance (i.e. the skip distance) up to which a particular frequency can be used. If we now write f'_{max} for the MUF and D_{skip} for the skip distance we have

$$f'_{max} = f_c \sqrt{\frac{L^2+4h_0^2}{4h_0^2}} \quad \ldots \quad \ldots \quad \ldots \quad (91)$$

and

$$D_{skip} = 2h_0 \sqrt{\frac{f'^2}{f_c^2} - 1}. \quad \ldots \quad \ldots \quad \ldots \quad (92)$$

(ii) FLAT EARTH—*Thick layer.*—Calculation of MUF when the reflecting layer is thick is necessarily complicated, because the ray follows an unknown curved path through the layer. Theoretical expressions may however be deduced if we assume some simple idealized variation of ionization with height. The case of parabolic gradient has been worked out and it can be shown that

$$D = y_m . x . \sin i_0 . \log_e \frac{1+x\cos i_0}{1-x\cos i_0} + 2h_0 \tan i_0 \quad \ldots \quad \ldots \quad (93)$$

where D is the transmission distance and $x = f'/f_c$ (see Fig. 54).

For a given value of x the distance D will be a minimum for the value of i_0 as given by

$$\frac{dD}{di_0} = y_m . x' . \log_e \frac{1+x'}{1-x'} - 2\tan^2 i_0 \left[y_m \cdot \frac{x'^2}{1-x'^2} - h_0 \right] + 2h_0 = 0,$$

or, $$2\tan^2 i_0 \left[\frac{y_m}{h_0}\cdot\frac{x'^2}{1-x'^2}-1\right]=\frac{y_m}{h_0}\cdot x'\log_e\frac{1+x'}{1-x'}+2, \quad \text{.. (94)}$$
where $x'=x\cos i_0$.

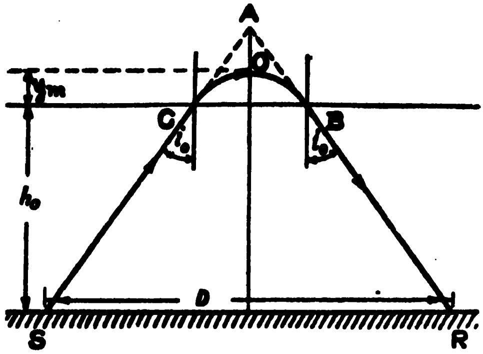


FIG. 54. Transmission path for the case of reflection from a thick ionized layer (flat earth).

The value of i_0 obtained from this equation evidently represents the minimum angle of incidence for a wave of frequency $f'(=xf_c)$ to be returned by the ionosphere. In other words f' is the maximum usable frequency for the angle of incidence i_0.

The distance D corresponding to the angle of incidence i_0, i.e., the distance for which f' is the maximum usable frequency under particular ionospheric conditions can now be obtained from Eq. (93). Obviously, this is the minimum distance D_{skip} of transmission for the frequency f', and to evaluate this, Eq. (93) shows that y_m and h_0 are required to be known separately; or, for a given ratio y_m/h_0 it will suffice to know the value of the quantity y_m+h_0. Curves may be drawn depicting the relation between D_{skip} and $x(=f'_{max}/f_c)$, the maximum usable frequency factor, with different values of y_m+h_0 ($=h_m$) for a series of values of the ratio y_m/h_0.

The above results may also be represented in a slightly different manner. The ratio f'_{max}/f_c may be plotted against the ratio y_m/h_0 with different values of y_m+h_0 for given distances of transmission. Complete sets of such curves have been drawn by the Radio Research Board, England, and are being used for computation of MUF for long distance radio communication.

(iii) CURVED EARTH—*Thin layer*.—If the earth is considered flat then it is obvious that as the angle of incidence of the wave is increased the skip distance will also increase until it extends to infinity. If, however, the curvature of the earth is taken into account, the skip distance cannot extend to infinity but attains a maximum value when the ray leaves the transmitter tangentially to the earth.

Let 2θ be the angle subtended at the centre of the earth by the transmission distance D (Fig. 55). We have

$$D=2R\theta, \quad \text{.. .. .. .. (95)}$$

and

$$\cos^2 i_0=\frac{(h_0+R-R\cos\theta)^2}{(h_0+R-R\cos\theta)^2+R^2\sin^2\theta}. \quad \text{.. .. (96)}$$

Under the condition of maximum D_{skip}, θ is a maximum and is given by

$$\cos\theta = \frac{R}{R+h_0} = 1 - \frac{h_0}{R}, \text{ since } \frac{h_0}{R} \ll 1.$$

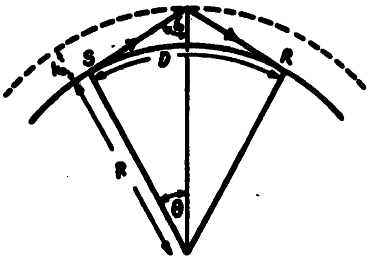


FIG. 55. Transmission over a curved earth by reflection from a thin ionized layer.

It is thus seen that θ is always very small. We may, therefore, write

$$\cos\theta = 1 - \frac{\theta^2_{max}}{2} = 1 - \frac{h_0}{R}$$

or, $$\theta^2_{max} = \frac{2h_0}{R}$$

and

$$[(D_{skip})_{max}]^2 = 4R^2\theta^2_{max} = 8Rh_0. \quad \ldots \quad \ldots \quad (97)$$

It can also be shown easily that

$$f'_{max} = f_c \sqrt{\frac{4\left(h_0 + \frac{D^2}{8R}\right)^2 + D^2}{4\left(h_0 + \frac{D^2}{8R}\right)^2}}. \quad \ldots \quad \ldots \quad (98)$$

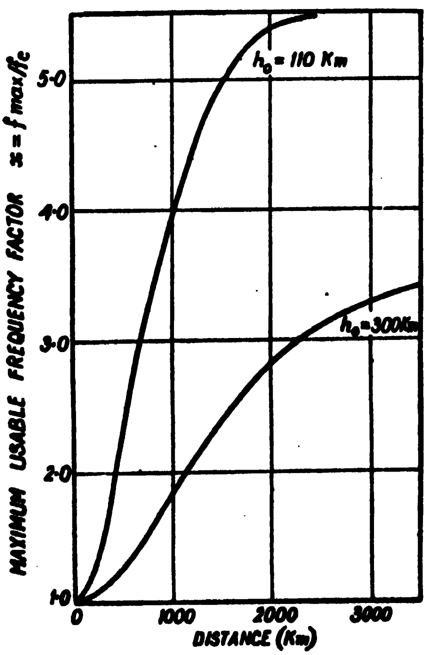


FIG. 56. Variation of maximum usable frequency factor with transmission range for curved earth and thin ionized layer.

In the case of a flat earth, R is infinite and this reduces to

$$f'_{max} = f_c \sqrt{\frac{L^2+4h_0^2}{4h_0^2}} \quad \ldots \quad \ldots \quad \ldots \quad (91)$$

Fig. 56 depicts the relation between f'_{max}/f_c and D as obtained from Eq. (98) for two typical values of h_0, *viz.*, 110 km., and 300 km., corresponding to Regions E and F_2 respectively.

(iv) CURVED EARTH—*Thick layer.*—In this case the transmission distance consists of two parts: D_1 corresponding to the path of the ray inside the ionosphere (which may be called the part-range) and D_2 corresponding to the path of the ray outside the ionosphere. This is shown in Fig. 57.

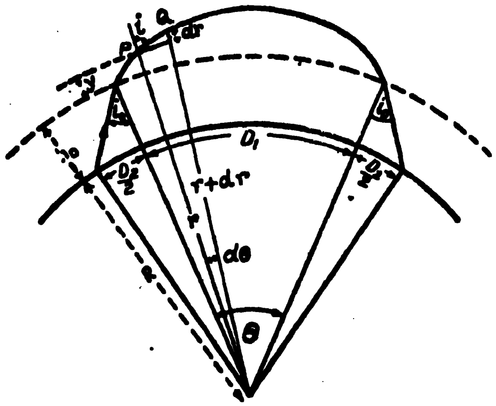


FIG. 57. Transmission over a curved earth by 'reflection' from a thick ionized layer. D_1, part-range, is the projection on the surface of the earth of the part of the ray path inside the ionized region.

It can be shown that the total range D is given by,

$$\frac{R}{R+h_0} \sin i_0 \,.\, x \,.\, y_m \,.\, \log_e \left[\frac{1-x^2 \dfrac{y_m}{R+h_0} \sin^2 i_0 + x \cos i_0}{1-x^2 \dfrac{y_m}{R+h_0} \sin^2 i_0 - x \cos i_0} \right]$$

$$+2R \cot i_0 - 2R \sqrt{\cot^2 i_0 - 2\frac{h_0}{R}} \quad \ldots \quad \ldots \quad (99)$$

It is convenient for practical purposes to have curves showing the relation between f'_{max}/f_c and D. For this purpose it is practicable to first use Eq. (94), which is appropriate for the case of a flat earth, for obtaining directly the value of i_0 for a given value of x. For this value of i_0, D_1 and D_2 are found from Eq. (99) remembering its first term is equal to D_1 and second term to D_2. x is then plotted against the total range D_1+D_2 as shown in Fig. 58, enabling one to read directly the maximum usable frequency. It is, of course, to be understood that the values of y_m and h_0 are known from the ionospheric data. The computation of i_0 from Eq. (94) for the case of a flat earth naturally involves some error. But it may be shown that this is not more than $2\frac{1}{2}$ per cent. This error may

be tolerated in view of the various simplifying assumptions made in the analysis.

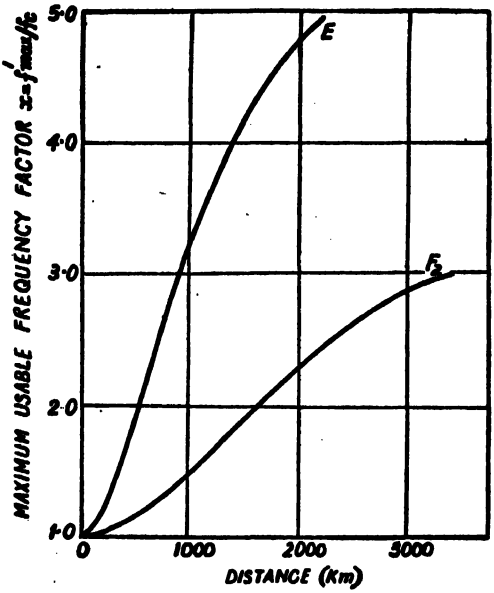


FIG. 58. Variation of the maximum usable frequency factor with transmission range for a curved earth and a thick ionosphere.

For very long distances, D_1 may have values as large as 1,200 km. and may include sunrise and sunset condition; the results cannot then be accurately applicable over the transmission path. The application of the result also becomes difficult for the case of very long distance transmission involving multihops.

It should be remembered that the computation of the theoretical curves is based on the assumption of a parabolic gradient of ionization in the region of maximum density. In determining the maximum usable frequency at a given time the values of y_m and h_0 (characterizing the parabolic distribution) used in the calculation must be those appropriate, for the region of the reflecting layer midway between the sending and the receiving stations at that time. This is rarely practicable. Further, for the purpose of computation one has to forecast the probable values of y_m and h_0 from past collected records over a sunspot cycle (*vide infra*).

(c) Smith's method (National Bureau of Standards, Washington)

We will now describe another method of determining the maximum usable frequency due to Newbern Smith [110] and used in the National Bureau of Standards, Washington, since 1936. The method is purely graphical and enables one to compute the MUF directly from the observed vertical incidence (P'—f) curves.

In describing this method we shall only indicate the principles on which it is based by considering the simplest case of a flat earth and no

magnetic field. For the more complicated cases in which the curvature of the earth and of the ionosphere cannot be neglected, the reader is referred to the original papers [109, 110].

We have already seen that in the case of a thin layer, the equation

$$\frac{f'^2}{f_c^2} = 1 + \frac{D^2}{4h_0^2} \quad \ldots \quad \ldots \quad \ldots \quad \ldots (100)$$

gives the relation between the maximum oblique incidence frequency f' and the distance of transmission D, where $f_c = \sqrt{(N_0e^2/\pi m)}$ and h_0 is the height of reflection of the ionized layer. In the case of a thick layer, however, the incident wave penetrates into the successive strata of increasing ionization and, therefore, one has to distinguish between the actual and the virtual heights of reflection.

It has been shown in Sec. 4(*e*) that if a wave incident vertically on the ionosphere is reflected from a point from which a wave at oblique incidence is also reflected, then the virtual heights reached by the two waves are the same, assuming a flat earth [69].

Let a wave of frequency f' incident at an angle i_0 be reflected from the point O in the ionosphere (Fig. 55) and returned to earth at a distance D from the transmitter. Let the virtual height be h_v.

Then we have as before

$$\tan i_0 = \frac{D}{2h_v}, \quad \ldots \quad \ldots \quad \ldots \quad \ldots (101)$$

$$\sin i_0 = \mu = \sqrt{1-\frac{Ne^2}{\pi m f'^2}} = \sqrt{1-\frac{f^2}{f'^2}}, \quad \ldots \quad \ldots (102)$$

and therefore,
$$\frac{D^2}{4h_v^2} = \frac{1-f^2/f'^2}{f^2/f'^2} = \frac{f'^2}{f^2} - 1, \quad \ldots \quad \ldots \quad \ldots (103)$$

where $f\left(=\sqrt{\frac{Ne^2}{\pi m}}\right)$

defines the value of electron number density at the actual height of reflection as distinct from that at the virtual height.

Transmission curves.—Let us now consider transmission over a given distance D with a wave of frequency f'. With the help of Eq. (103) we may draw a curve depicting the relation between h_v and f for these values of f' and D. Any point (h_v, f) on this curve means that a wave of frequency f' incident obliquely on the ionosphere will be transmitted over distance D if its virtual height of reflection is h_v and if the ionization distribution in the ionosphere is such that a wave of frequency f incident vertically would be reflected from the same virtual height h_v. This frequency is called the *equivalent normal incidence frequency* and the $(h_v—f)$ curve is called a *transmission curve*, as it depicts conditions necessary for transmission for the given distance and frequency.

Let us now consider an experimental (h'—f) curve observed at normal incidence. A point (h', f) on this curve means that a frequency f is reflected back normally from the virtual height h'. If the transmission curve for the given distance D and frequency f' is such that for the same value of virtual height ($h_v = h'$) the equivalent normal incidence frequency happens to be equal to f, then and then only can the wave of frequency f' be reflected at the height h_v and transmission over the distance D is possible. In other words, if the transmission curve for a frequency f' and distance D and the observed (h'—f) curve have a common point, it signifies that transmission is possible under the existing conditions of the ionosphere with a wave frequency f' over a distance D. If the (h'—f) curve observed at a given time and the (h_v—f) curve do not intersect at any point it shows that the ionospheric conditions at that time are such that transmission over a distance D on a wave of frequency f' is not possible. Let us now draw a number of curves for increasing values of f' and a given D and superimpose them on the observed (h'—f) curve. It will be seen that up to a certain value of f', the (h_v—f) curves intersect the (h'—f) curve, but beyond this value the (h_v—f) curves pass away from it. Evidently then, this particular value of f' is the maximum usable frequency for transmission over the distance D under ionospheric conditions depicted by the observed (h'—f) curves.

Approximate practical method of obtaining the maximum usable frequency—The above method of finding the maximum usable frequency is obviously very laborious. A simpler and quicker method may be adopted as follows for the case under consideration (no magnetic field, flat earth). We have the equation

$$f = f' \cos i_0. \quad .. \quad .. \quad .. \quad .. \quad (104)$$

Instead of drawing the (h_v—f) curve for a given f' and D we may plot h_v against $\cos i_0$ for a given D from Eq. (101). Since for a given f', f bears a constant relation with $\cos i_0$ such a curve may be made to represent the transmission curve for the distance D for a particular frequency f' provided that to each value of $\cos i_0$ we assign a value of f appropriate to f' given by Eq. (104). Alternatively, we may easily ascertain to what frequency f' the curve pertains if the value of f corresponding to one particular value of $\cos i_0$ is known.

This principle is utilized for rapid calculation of the maximum usable frequency. The procedure may be outlined as follows.

Let us take the two curves on identical semi-logarithmic paper (Fig. 59):

(1) An observed (h'—f) curve with f plotted logarithmically and h' linearly.

(2) An (h_v–$\cos i_0$) curve, satisfying Eq. (101) for a given value of D with $\cos i_0$ plotted logarithmically and h_v linearly.

Note that h_v and h' are plotted in the same scale.

Superimpose curve (2) on curve (1) taking care that points of equal ordinates of the two curves coincide, that is, the h_v-axis falls on the h'-axis with their origins coinciding.

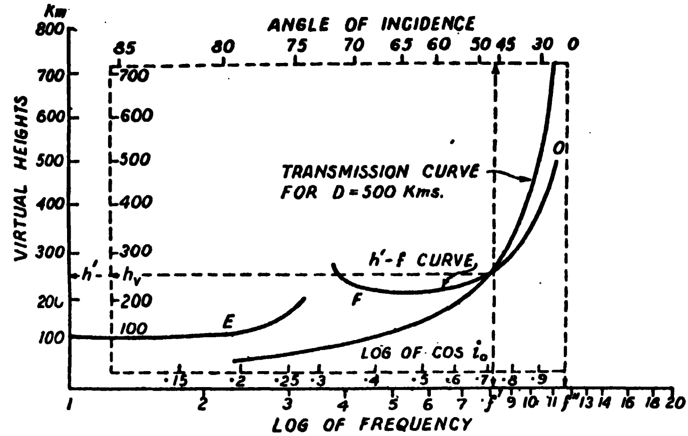


FIG. 59. Illustrating the graphical method of determination of the maximum usable frequency.

Let us suppose that the two curves cut at the point h_1, f_1 on the $(h'—f)$ curve. Let the co-ordinates of this point on the $(h_v—\cos\, i_0)$ curve be h_1, $\cos\, i_1$. If we now let the $(h_v—\cos\, i_0)$ curve represent the transmission curve for an oblique frequency f' such that f_1 is equal to $f' \cos\, i_1$ then since the abscissae of both the curves are plotted logarithmically on the same paper, the abscissae of the $(h'—f)$ curve will exactly represent also the abscissae (equivalent vertical incidence frequencies) of this transmission curve.

Let us now shift the $(h_v—\cos\, i_0)$ curve towards increasing f so that it just touches the $(h'—f)$ curve. Let this common point of contact be h', f on the $(h'—f)$ curve and h_v, $\cos\, i_0$ on the $(h_v—\cos\, i_0)$ curve. This latter curve will now represent the transmission curve for the distance D for a new frequency f' equal to $f/\cos\, i_0$; as before, the f-co-ordinates will be common to both the $(h'—f)$ curve and the transmission curve. A further shift of the $(h_v—\cos\, i_0)$ curve towards the right would take it away from the $(h'—f)$ curve. Thus $f' = f/\cos\, i_0$ represents the maximum frequency for oblique transmission over the distance D. Obviously f' is the value of f on the f-axis corresponding to $\cos\, i_0 = 1$.

The method of practically determining the maximum usable frequency for transmission over a given distance D may thus be summarized as follows.—

(i) Plot the virtual height-frequency $(h'—f)$ curve with data obtained from vertical incidence experiments on semi-logarithmic paper, h' being plotted on the vertical linear axis and f on the horizontal logarithmic axis.

(ii) Plot $(h_v—\cos\, i_0)$ curve for the equation

$$h_v = \frac{D}{2} \cot i_0,$$

or,

$$\cos i_0 = \frac{h_v}{\sqrt{h_v^2 + \left(\frac{D}{2}\right)^2}},$$

h_v being plotted on the vertical linear axis and $\cos i_0$ on the horizontal logarithmic axis.

The two curves are to be plotted on the same type of paper.

(iii) Superimpose curve (ii) on curve (i) so that the h_v-axis falls on th h'-axis with their origins coinciding. Shift (h_v—$\cos i_0$) curve horizontally until it just touches the (h'—f) curve.

(iv) Read the frequency f on the horizontal axis corresponding to $\cos i_0 = 1$. This frequency is the maximum usable frequency for transmission over the distance D under ionospheric conditions represented by the observed (h'—f) curve.

(*d*) Ionospheric Predictions

For the proper maintenance of world radio communication it is very necessary to know well in advance the *Maximum Usable Frequencies* for the different ionospheric layers (*via* which the transmission may take place) between the two distant communicating points on the earth. Such MUF's may be calculated by any one of the methods described above if one is able to predict beforehand the ionospheric characterstics of the corresponding layers for vertical incidence. Methods have been devised for making such predictions and are briefly described below:

(i) *Predictions for E and F_1 regions.*—For short distance (that is, up to 2000 km.), and during day-light, transmission generally takes place *via* E-layer and occasionally *via* F_1-layer. Predictions for these layers are comparatively easy and accurate on account of the regularity with which the average values of the characteristics of these layers vary with the hour of the day, the season of the year and also with the solar cycle. The first step is to predict the characteristics for vertical incidence at the point midway between the transmitting and the receiving station. This is done as follows:

The probable average noon-time value for the month of prediction is first determined from the history of the layer for the previous solar cycles. Next, from this noon-time value, the diurnal variation is predicted knowing the trend of such variation for the particular season and for the relevant phase of the solar cycle. This latter is obtained from representative variation curves drawn on the basis of past observations [112]. Fig. 60 gives after Naismith, the predicted noon values of f^0E, and f^0F_1 at Slough, England.

Since the characteristics of these regions closely follow the $\sqrt{\cos \chi}$ law, world contour charts (*vide infra*) for the penetration frequencies are easily constructed. From the values obtained from these charts and from

the corresponding MUF factors, the maximum usable frequencies may be computed.

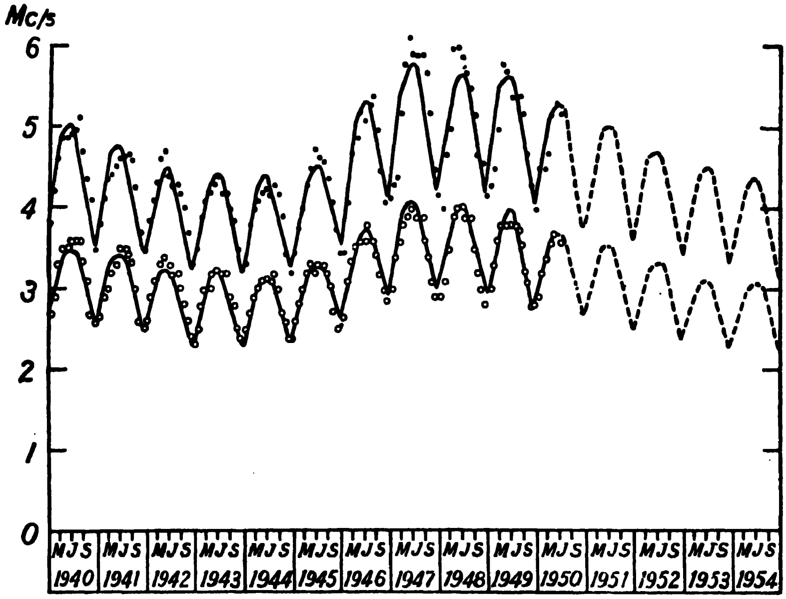


FIG. 60. Calculated variation of monthly average midday critical frequencies, f^0F_1 and f^0E for the period 1940–54. The data are for Slough, England. The continuous portions of the curves are based on observed sunspot data and the broken line portions on predicted values. (After Naismith.)

• —observed monthly average midday f^0F_1

∘ —observed monthly average midday f^0E

(ii) *Predictions for F_2-region.*—Most of the long distance communications are carried out *via* E_s and F_2 layers. Predictions for the former are difficult and are more or less empirical. Predictions for the latter are also difficult on account of the irregular nature of the diurnal, seasonal and solar cycle variation of its characteristics. Nevertheless, predictions have been made possible by the availability of the records from a fairly large number of stations (54 in 1950) established during World War II and still continuing their observations in the different parts of the world.

The vertical incidence characteristics (i.e., the critical frequencies and the MUF factor, which latter is a function of y_m and h_0) at the predicting station, may be obtained in advance either graphically, as is done at the National Physical Laboratory, England, or computed as at the National Bureau of Standards, Washington.

At the former, the monthly mean values for a particular hour or the monthly averages of 24 hours are computed from past data and a graph is drawn showing the variation of each of these values from one month to another for a complete solar cycle. Monthly average values for the solar cycle under consideration are plotted in a similar graph. From the trend

of the curve of the previous solar cycle, the second graph is extrapolated and predictions are made three months in advance.

In the method adopted by the National Bureau of Standards, Washington, the annual running average value of Region F_2 critical frequency for each hour is computed, and the mean trend from year to year is determined as a function of the Wolf number (see Appendix, Sec. 8). From the yearly values so obtained seasonal values are computed with the help of seasonal coefficients calculated from past data. Finally, the monthly mean values are obtained from the seasonal means with the help of the monthly coefficients calculated also from past data.

(iii) *World Contour Charts.*—For the purpose of world prediction it is necessary to know the f^0F_2 and the MUF factors in advance for points all over the world. As already indicated, this has now been made possible, thanks to the establishment of a number of stations in the different parts of the world. The representative data for the various latitudes and longitudes are obtained by extrapolation from the recorded data at these stations [113].

However, as the F_2 characteristics show marked longitude effect (see Sec. 11) it is not possible to carry out extrapolations over long distances from the observing stations. Hence, predictions are made in sections for limited zones of the earth (see, however Ref. [113a]). It has been the convention to divide the world into zones with reference to the meridian passing through the geographic and geomagnetic poles and also two intermediate zones contained within longitudes $\pm 60°$ of the same. The predicted world values of the vertical incidence characteristics, the f^0F_2 and the MUF factors, are thus depicted in the form of contour charts for four zones: East Zone (50°E. to 170°E.), Intermediate Zone (170°E. to 130°W.), West Zone (130°W. to 10°W.) and second Intermediate Zone (10°W. to 50°E.).

From the corresponding values of f^0F_2 and the MUF factors it is now easy by simple multiplication to obtain the MUF's. The usual practice is to construct MUF charts for transmission distances over 4000 km. and to use them as follows for greater distances.

The MUF's for the control points at distances 2000 km. from the sender and the receiver are read from the chart (*vide infra*). The lower of the two values is taken as the MUF for the distance concerned. For distances less than 4000 km. the MUF factor chart for 3000 km. is utilized along with the f^0F_2 charts. A set of curves representing the variation of MUF factor with transmission distance is drawn once for all.

Fig. 61 is an example of World Contour Chart, of MUF (4000 km.) prediction for the East Zone, for September, 1950, as prepared by the Radio Research Station, Slough, England. The following example will illustrate the use of the chart (for transmission distance over 4000 km.). Suppose that it is required to determine the MUF for transmission from Calcutta (22·5°N., 88·5°E.) to Cairo (30°N., 31°E.) at 0600 hours local time (Calcutta

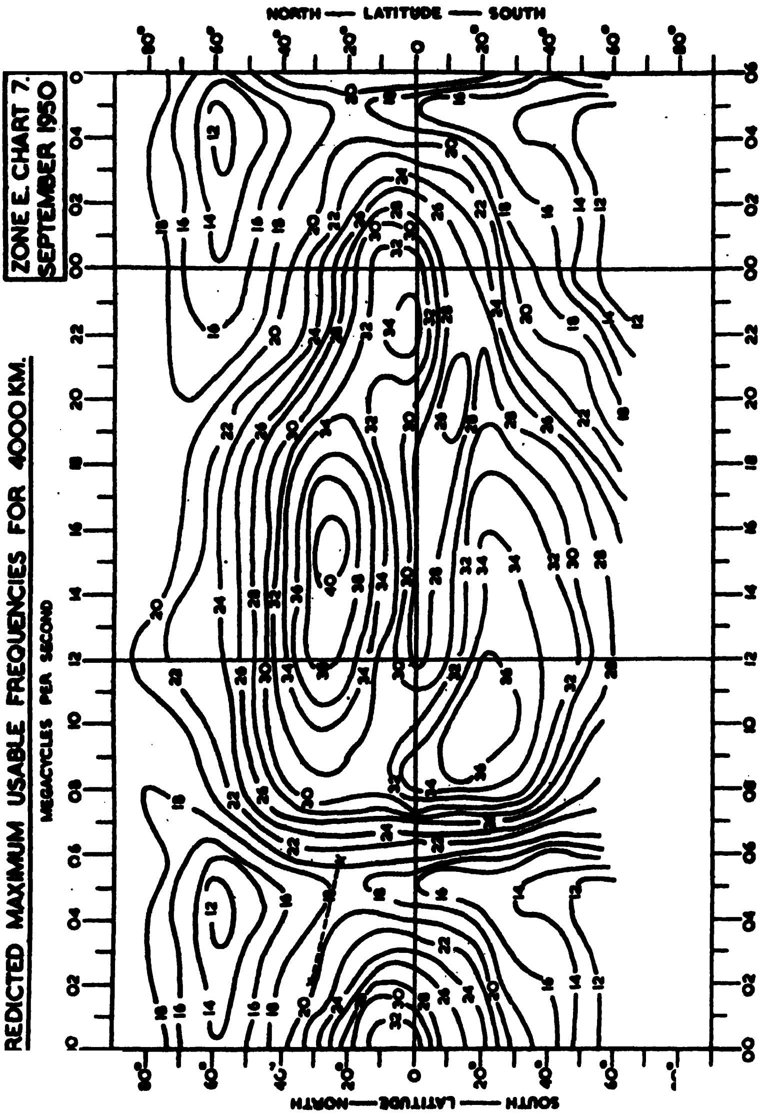


FIG. 61. World Contour Chart (East Zone) of predicted maximum usable frequency for transmission distance 4000 km. for September 1950 as prepared by the Radio Research Station, Slough, England.

for the above month in 1950. We first determine, the latitude and longitude of the midpoint on the great circle joining Calcutta and Cairo. These are 25·5°N., 55·5°E. From the longitude we find that when it is 0600 hours local time at Calcutta, it is 0406 hours local time at the midpoint. From the chart we find that at the point where latitude is 25·5°N. and local time is 0406 hours, the MUF is 19 Mc/s. This is, therefore, the MUF for transmission from Calcutta to Cairo at 0600 hours local mean time during September, 1950.

The National Physical Laboratory, England, publishes nine World Contour Charts of prediction for each month, three for f^0F_2, three for the

MUF factors for 3000 km. transmission and three for the MUF's for 4000 km. transmission.

It may be noted that the MUF calculated as above may be inapplicable owing to the presence of ionospheric irregularities (*vide infra*). These irregularities are associated with certain geomagnetic and solar disturbances which can be predicted from previous data. It is therefore customary for the various ionospheric laboratories (e.g. the Central Radio Propagation Laboratories, National Bureau of Standards, Washington) to report the geomagnetic and solar indices along with the ionospheric data. [See Chap. VII, Sec. 2(*a*).]

(iv) *Prediction of MUF from observation of long distance scatters.*—Observations on long-distance scatter (or ground scatter) (*vide* Sec. 13) may also be utilized for determining the MUF for long distance transmission [113*b*]. The MUF's predicted from such observations are free from uncertainties due to unpredictable ionospheric changes. These predictions are short-term predictions (24 hours) but, are much more reliable than the long-term ones.

10. ORIGIN AND STRATIFICATION OF THE IONOSPHERE

(*a*) Introduction

The process by which an ionized stratum is formed by action of solar ultraviolet radiation may be understood as follows. Consider the earth to be enveloped by a single gaseous element the density of which decreases exponentially upwards. If a beam of monochromatic radiation enters the atmosphere from outside and is absorbed in producing ionization, then the rate of ion production will be a maximum at a certain level. This is because the rate of absorption at any height is controlled by two factors—the intensity of the incident radiation and the density of the absorbing gas. As the radiation enters the atmosphere from above, the former decreases and the latter increases. The two opposing factors thus combine to produce a maximum at a level which is determined on the one hand by the coefficient of absorption of the gas for the particular radiation and on the other by the rate of decrease of the density upwards. This idealized process of layer formation has been studied in detail by Chapman [114] and will be discussed in the next section. The method of attack is simple and direct, and yields valuable information regarding the expected nature of the diurnal and seasonal variations of the ionization, i.e., the variations caused by the changing inclination of the ionizing solar rays.

An earlier attack on the problem was made by Pannekoek [115] who started with Saha's theory [116] of thermal ionization as modified by Woltjer [117] and Milne [118]. This method takes into account the continuous nature of the solar radiation and has been successfully adapted by Bhar [119] for explaining the ionospheric stratification. Pannekoek's method and its extension will be discussed later in this section.

(b) Formation and properties of a simple Chapman region

Consider a beam of monochromatic radiation of unit cross-section passing through the atmospheric layers at an inclination χ to the vertical. Let the absorption coefficient per unit mass of the gas be A and the intensity of the incident radiation be I at a height h above the ground. Then, if χ is not greater than 85°, the change of intensity dI after the radiation has passed through a layer of thickness dh at the height h will be given by

$$dI = AI\,dh \sec\chi\,\rho_0 \exp(-h/H) \qquad \text{(105)}$$

where ρ_0 is the density of the gas at the ground level and $H(=kT/mg)$ is the 'scale height'.

On integration,

$$I = I_0 \exp\{-A\rho_0 H \sec\chi \exp(-h/H)\} \qquad \text{(106)}$$

where I_0 is the intensity of the radiation before it enters the atmosphere. If β be the number of ions produced by absorption of unit quantity of radiation, then the rate of production of ions per cm.^{3} at the height h is given by

$$q = \beta \frac{dI}{dh} \cos\chi \qquad \text{(107)}$$

$$= \beta A I_0 \rho_0 \exp\{-h/H - A\rho_0 H \sec\chi \exp(-h/H)\}. \qquad \text{(108)}$$

On differentiation, it is easily seen that q will be a maximum at the height $h(\chi)$ given by

$$\exp\frac{h(\chi)}{H} = A\rho_0 H \sec\chi \qquad \text{(109)}$$

and the maximum rate of ion production will be

$$q(\chi) = \beta I_0 \cos\chi/H \exp 1. \qquad \text{(110)}$$

When $\chi = 0$, Eqs. (109) and (110) reduce to

$$h_M = H \log_e A\rho_0 H \qquad \text{(111)}$$

and

$$q_0 = \beta I_0/H \exp 1. \qquad \text{(112)}$$

We may express Eq. (108) in terms of h_M and q_0 which are the values of $h(\chi)$ and $q(\chi)$ for vertical incidence. Thus

$$q = q_0 \exp\left[\frac{h_M + H - h}{H} - \sec\chi \exp\frac{h_M - h}{H}\right] \qquad \text{(113)}$$

Let us reckon heights from h_M as the reference level and in terms of H as unit. Then we may write Eq. (113) as

$$q = q_0 \exp\{1 - z - \sec\chi \exp(-z)\} \qquad \text{(114)}$$

where $z = \dfrac{h - h_M}{H}$.

As mentioned before this equation is valid only for values of χ less than 85°. Chapman has extended this analysis to cover also the case of

grazing incidence. However, in middle latitudes near the level of maximum ion production the results derived from the above simple formula need very little correction.

We now note a few important results derived from the above analysis.

(i) *Variation of electron number density with zenith distance of the sun under quasi-equilibrium condition.*—For the ionospheric regions the electron (or ion) number density changes slowly with time during full daylight hours. This means that at any instant in daylight hours the rate of ion production (q) and the rate of ion decay approximately balance each other. If N_e be the electron (or positive ion) number density at any instant t at height h, then, assuming that recombination is the only process of decay, the number of electrons lost per cm.^{3} per second is αN_e^2. Hence for such quasi-equilibrium condition we put

$$\frac{dN_e}{dt} = q - \alpha N_e^2 \fallingdotseq 0 \qquad \ldots \qquad \ldots \qquad \ldots (115)$$

$$\text{or,} \quad q \fallingdotseq \alpha N_e^2 \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots (116)$$

Substituting the value of q from Eq. (110) we get

$$\alpha N_e^2 = \beta I_0 \cos\chi / H \exp 1,$$

or

$$N_e = \text{const.} (\cos\chi)^{\frac{1}{2}}.$$

We thus arrive at the important result that the electron number density at any latitude, under equilibrium condition, is proportional to the square root of the cosine of the zenith angle of the sun.

The case when equilibrium condition does not hold, e.g. during sunrise and sunset periods will be discussed in sub-section (v) below.

(ii) *Particle concentration of the active constituent at the height of the maximum of the layer.*—$\rho(X)$, the density of the active gas (X) at any height is given by

$$\rho(X) = \rho_0 \exp\left(-\frac{h}{H}\right).$$

Then, from Eq. (109)

$$\rho(X) = \frac{\cos\chi}{AH}. \qquad \ldots \qquad \ldots \qquad \ldots \qquad (116.1)$$

This equation gives the particle concentration $\rho(X)$ of the active constituent X at the height of the maximum of the layer.

(iii) *Distribution of electron number density with height.*—The distribution of electron number density with height under equilibrium conditions may be obtained as follows. We have from Eqs. (114) and (116), at the level z for a zenith angle χ,

$$N_e = N_0 \exp \tfrac{1}{2}\{1 - z - \sec\chi \exp(-z)\},$$

where N_0 $(= \sqrt{q_0/\alpha})$ is the electron number density at the level of maximum ionization for $\chi = 0$.

Now, if χ is not large, i.e. when the sun is not far from the zenith, $\sec \chi = 1$, and we have for the region for which z is small, approximately

$$N_e = N_0\left(1-\frac{z^2}{4}\right)$$
$$= N_0\left(1-\frac{y^2}{4H^2}\right) \quad \ldots \quad \ldots \quad \ldots (117)$$

where $y\ (= h-h_M)$ is the height measured from h_M, the level of N_0.

Eq. (117) is that of a parabola of the form shown in Fig. 62.

We thus see that in the neighbourhood of the level of maximum ionization the distribution of electron number density with height is roughly parabolic. It may be shown that this parabolic distribution holds good to an accuracy of 95 per cent within the region $h_M \pm H$. Such a region may be called a simple Chapman region.

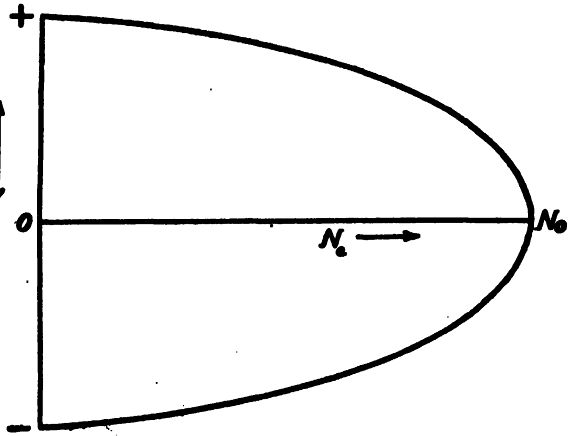


FIG. 62. Calculated parabolic distribution of electron number density in an ionospheric layer. The actual distribution approximates to this curve to within 5 per cent up to a distance of H km. above and below the level of maximum ionization.

(iv) *Absorption of radio waves caused by a simple Chapman region.*— It is important to calculate the total absorption suffered by a radio wave in its passage through a simple Chapman region in which it is not deviated. This can be done as follows [120]. It may be shown from Eq. (37.1) that for the ordinary wave the absorption coefficient for an underlying non-deviating region is given by

$$k = \frac{2\pi e^2}{mc}\cdot\frac{N_e\nu}{\nu^2+(p+|p_L|)^2}\cdot \quad \ldots \quad \ldots \quad \ldots (118)$$

If ν is assumed to vary with height according to the simple exponential law $\nu = \nu_0 e^{-z}$, and if the variation of N_e be that for a simple Chapman region, k is given by

$$k = \frac{2\pi e^2}{mc}\cdot N_0\nu_0\left[\frac{\exp \frac{1}{2}\{1-3z-\sec\chi\exp(-z)\}}{\nu_0^2 e^{-2z}+(p+|p_L|)^2}\right]. \quad \ldots (119)$$

Now, consider a wave reflected from the F-region say, which in its up and down journey has passed twice through an underlying absorbing, non-deviating Chapman layer. The apparent reflection coefficient (ρ) of the F-region is then given by

$$-\log \rho = 2\int k\,dz$$

$$= \frac{4\pi e^2 N_0 H}{mc}\left[\frac{\exp 1}{\nu_0(p+|p_L|)}\right]^{\frac{1}{2}} X \quad \ldots \quad (120)$$

where

$$X = \int_0^\infty \frac{t^{1/2}}{1+t^2}\, e^{-at}\, dt, \quad \ldots \quad (121)$$

$$a = \frac{(p+|p_L|)\sec\chi}{2\nu_0}. \quad \ldots \quad (122)$$

Now, for reflection from Region F, a sufficiently good approximation is $p \gg \nu$ so that $a \gg 1$ and $X = \frac{1}{2}\sqrt{\pi/a^3}$. With these approximations and substituting for X in (120), we have

$$-\log \rho = 4{\cdot}13\left(\frac{4\pi e^2}{mc}\right)\frac{N_0 \nu_0 H}{(p+|p_L|)^2}\cdot \cos^{3/2}\chi. \quad \ldots \quad (123)$$

It is thus seen that $|\log \rho|$ varies as $\cos^{3/2}\chi$. Experimental results show that this law is obeyed only approximately [120, 45].

(v) *Variation of ionization with time and height (general expression).*— In sub-sections (i) and (ii) above we obtained simple relations for the variations of the ionization density (N_e) with the hour of the day and with height under certain restricted conditions. For the former, a quasi-equilibrium condition was assumed to hold throughout the day-light hours; for the latter we considered only the region near the maximum ionization density. For closer study, when the lack of equilibrium condition has to be taken into account, ionization variations can be obtained as follows in terms of q_0 and N_0, the rates of electron production and electron decay at the equinoxes over the equator, when and where the equilibrium condition may be assumed to hold more strictly.

We have

$$q_0 = \alpha N_0^2 \text{ or } N_0 = (q_0/\alpha)^{\frac{1}{2}}. \quad \ldots \quad (124)$$

Substituting for q from Eq. (114) in Eq. (115) we have

$$\frac{dN_e}{dt} = q_0 \exp\{1-z-\sec\chi \exp(-z)\} - \alpha N_e^2. \quad \ldots \quad (125)$$

If ϕ be the local time in radians and t is in seconds, then

$$t/86400 = \phi/2\pi. \quad \ldots \quad (126)$$

We express N_e in term of N_0 and write

$$k = \frac{N_e}{N_0}. \quad \ldots \quad (127)$$

We also write

$$\sigma_0 = 1/(1{\cdot}37\times 10^4 N_0\alpha). \quad .. \quad .. \quad .. \quad (128)$$

Substituting for N_e and t in Eq. (125) we have

$$\sigma_0 \frac{dk}{d\phi}+k^2 = \exp[1-z-\sec\chi\exp(-z)]. \quad .. \quad .. \quad (129)$$

The above expression, however, does not hold good for grazing incidence ($\chi>85°$). For such case we replace after Chapman [114] sec χ by a somewhat complicated function $f(R+z, \chi)$ where $R=(a+h_0)/H$ which approximates to sec χ when $\chi<85°$.

During the hours of darkness, the right-hand side must be zero, i.e.,

$$\sigma_0 \frac{dk}{d\phi}+k^2 = 0. \quad .. \quad .. \quad .. \quad .. \quad (130)$$

On integration,

$$k = \frac{\sigma_0}{\phi+C}, \quad .. \quad .. \quad .. \quad .. \quad (131)$$

where C is an integration constant.

If k_r and k_s are the values of k at sunrise and sunset, we have

$$\frac{1}{k_r}-\frac{1}{k_s} = \frac{\phi_r-\phi_s}{\sigma_0}, \quad .. \quad .. \quad .. \quad (132)$$

where ϕ_r and ϕ_s are the local times for sunrise and sunset.

Typical curves may now be drawn to illustrate how according to Chapman's theory of layer production the ionization density varies with

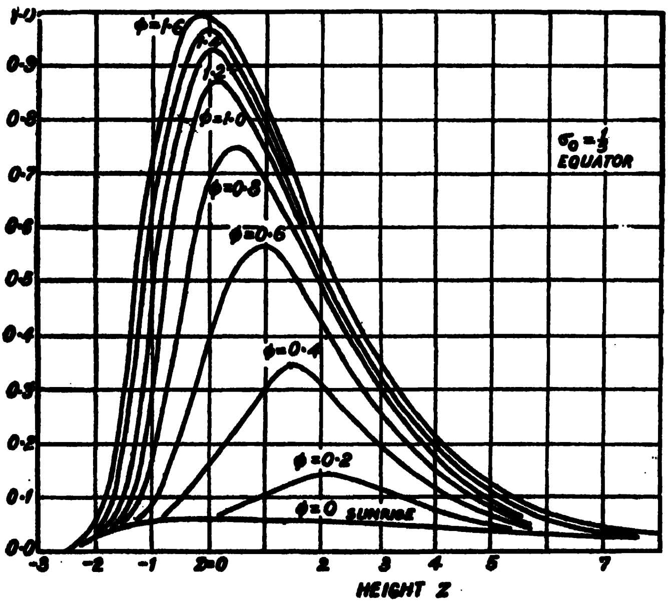


FIG. 63. Distribution of electron number density with height at different hours of the day at the equator during the equinoxes, for $\sigma_0 = 0{\cdot}2$. The ordinates represent $k = N_e/N_0$. (After Chapman.)

height (for different values of χ) and also how it varies throughout the 24 hours of day and night (in different seasons). For this, we note that in the expression for $1/\sigma_0$, N_0 is of the order 10^5/cm.^{3} and α is of the order 10^{-9} cm.^{3}/s. Thus σ_0 is of the order unity. Illustrative curves may therefore be drawn for the probable typical values of σ_0. In Fig. 63 the height variation of k i.e., of N_e ($= N_0 k$) is shown after Chapman for $\sigma = 0{\cdot}2$, for different values of ϕ i.e. for the different hours of the day. In Fig. 64 the hourly variation of k is depicted for the latitude of Calcutta for five days in different seasons.

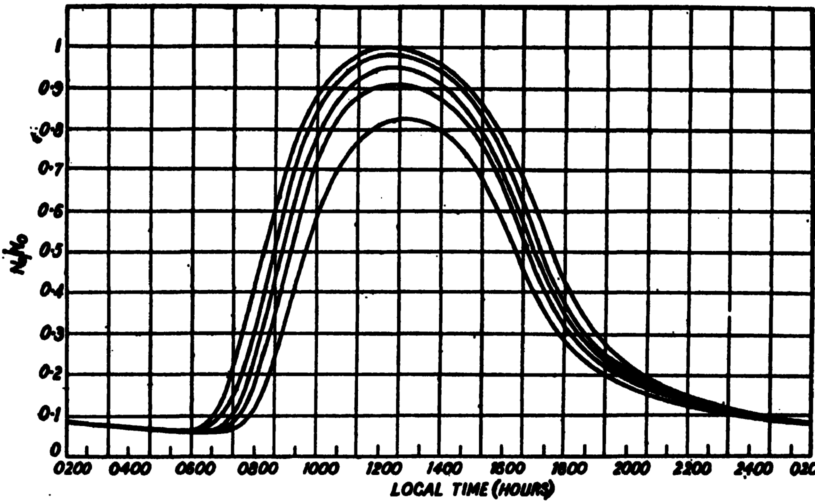


FIG. 64. Hourly variation of ionization for five selected days for the latitude of Calcutta for $\sigma_0 = 0{\cdot}25$ (after Chapman's theory). Starting from the top the curves successively correspond to (i) Summer solstice, (ii) δ (sun's declination) $= +10°$, (iii) Equinoxes, (iv) $\delta = -10°$, and (v) Winter solstice.

In Sec. 11 we shall discuss how far the observed variations of ionization in the different ionospheric regions follow these ideal curves.

(vi) *Effects of departures from ideal Chapman conditions.*—As already remarked the Chapman layer is formed under highly idealized conditions viz., the atmosphere is isothermal (or, rather $H = kT/mg$ is constant), the ionizing radiation is monochromatic and the recombination coefficient is constant with height. None of these conditions is satisfied in practice and the departures distort the ideal layer and make it diffuse. Theoretical investigations have been carried out to determine the various effects of the departures from the ideal conditions. Nicolet and Bossy [121], for example, have examined how the maximum ionization density and the reflection coefficient vary with the changing zenith distance of the sun when H is not constant but varies with height in some specified manner. Fig. 65 depicts after Gledhill and Szendrei [121a] how the ionization density will vary with height when there is a rising temperature gradient. It will be noticed that when the gradient is large there is no layer

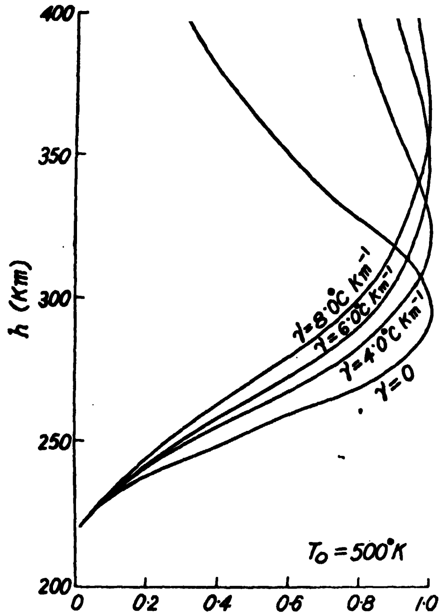


FIG. 65. Curves illustrating distribution of ionization (N_e/N_0) with height for the F_2-layer when the temperature increases with height.
T_0—temperature at the base of the layer (datum level).
γ—temperature gradient. (After Gledhill and Szendrei.)

formation. These results are of importance in interpreting the observed data and determining the probable profiles of the ionized layer.

(c) Pannekoek's method of computing upper atmospheric ionization

The above analysis of the ionization produced by monochromatic radiation may not represent the actual conditions since the incident radiation is not monochromatic being that due to the sun. The method of calculation taking into account the continuous nature of the solar radiation was first given by Pannekoek [115] in 1926. It will be seen, however, that the fact that the radiation is continuous and not monochromatic does not greatly affect the results of Chapman's analysis. This is because, for a particular process of ionization, only a short range of wavelength near the critical wavelength is effective in producing ionization in appreciable amount. In fact, the totality of the effective radiations, which lie within a small range adjacent to the critical ionizing wavelength, may be regarded as a quasi-monochromatic radiation having a finite width. Thus, the methods of attack developed by Chapman and by Pannekoek yield practically the same result regarding the formation of ionized layers [122].

In calculating the ionization of the earth's atmosphere Pannekoek made use of an analysis carried out by Woltjer [117] and Milne [118] in connection with the ionization of stellar atmospheres.

Pannekoek's method, as was pointed out by Mitra [123], may be successfully adapted to explain the origins of the various ionospheric 'layers'.

Bhar [119] in particular has carried out detailed calculations on layer formations by this method and has shown that

$$\frac{\eta^2}{1-\eta^2} p = K \exp\left(-\gamma n_s/\theta\right) \quad .. \quad .. \quad .. \quad (133)$$

where η—fraction of gas ionized,

p—total pressure due to the ions, electrons and neutral atoms (dynes/cm.^{2}),

n_s—number of molecules or atoms/cm.^{3} at height z above a certain datum level ($= n_0 e^{-\theta z}$, where n_0 is the number of molecules/cm.^{3} at the datum level),

θ—a constant determining the nature of the distribution of density with height of the gas,

K—a constant depending on the effective band of ionizing radiation and the constituent ionized, and is given by

$$K = \frac{w}{4\pi} \frac{(2\pi m)^{\frac{3}{2}}}{h^3} k^{\frac{5}{2}} \cdot T_1{}^3 T^{-\frac{1}{2}} e^{-\Omega/kT} \frac{\int_{x_{1_0}}^{x_1{}'} x_1{}^2 e^{-x_1} dx_1}{\int_{x_0}^{x'} x^2 e^{-x} dx} \quad \ldots (134)$$

where Ω—energy of ionization/atom,

T_1—temperature of the sun in degrees absolute considered as a black body,

T—temperature of the gas in degrees absolute,

w—solid angle subtended at the earth by the sun,

$x_1 = h\nu/kT_1$, $x = h\nu/kT$. The integrations are carried out between limits ν and ν'.

With the help of the above equation Bhar has calculated the distributions of electron concentration with height due to the ionizations of O, N_2 and O_2 by making plausible assumptions regarding the distributions and absorption coefficients of these gases. These are discussed in the following section.

(*d*) Production of the different ionospheric layers

With the solitary exception of sporadic E, the other layers are known to owe their origin to the ionizing action of the solar ultraviolet radiation. In what follows the various theories proposed for the production of the different layers will be briefly discussed.

(i) *Region E.*—Production of an ionized layer at the observed height of the E-layer was for a long time a puzzle. It was, however, pointed out by Mitra [123] and by Bhar [119] (see sub-sec. (*c*)) and also independently by Wulf and Deming [124] that the E-layer may be formed by ionization of O_2 in the region of transition where molecular oxygen is dissociated and where, as a consequence, its density falls off with height much more rapidly than for a simple isothermal atmosphere. Bhar assumed that the ionization was at the second ionization potential of O_2 and calculated by

the Pannekoek method (as modified by him, *vide supra*) the ionization density distribution with height. (See Fig. 66.) In contrast to Bhar's assumption Wulf and Deming prefer ionization at the first ionization potential to be effective. Objections to both these assumptions have, however, been raised. It has been pointed out that the ionizing band λλ 770–661 at the second ionization potential lies within the region of strong absorption by atomic oxygen and, as such, will be used up before it reaches the transition level. Again, for the first ionization potential the absorption cross-section is so low that no ionization maximum would be produced at the required level. (Ionization at the first ionization potential of O_2 is now believed to produce the *D*-region, *vide infra*.)

Two different theories have been proposed to meet the difficulties. According to one, emission from the solar corona is supposed to provide

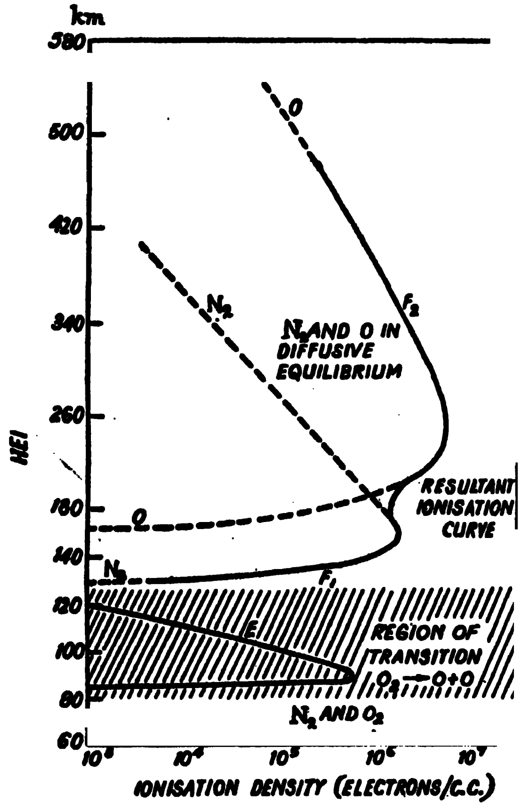


FIG. 66. Curves illustrating calculation of ionization distribution with height by Pannekoek's method (as modified by Bhar). This method is applicable when the ionizing radiation instead of being monochromatic is continuous in wavelength. It is to be noted that the *E*-layer is formed in the region where the density of O_2 decreases rapidly with height due to photo-dissociation. The F_1-layer is shown as due to ionization of N_2, and the F_2-region as due to ionization of O. According to contemporary ideas, however, F_1-region is produced by photo-ionization of atomic oxygen and F_2-region by a sort of bifurcation of the F_1-region.

the ionizing radiation [128]. It is shown that if the sun emits photon groups at about 325A or at about 1300A then an ionized layer could be produced at the required level of Region E. Of the two groups, the 325A group appears to be the more likely agent. According to the other, due to Nicolet [125], attention is directed to the existence of a number of strong absorption bands, superimposed on the feeble absorption continuum of O_2 in the energy range 12·20 to 13·55 eV. According to Nicolet, absorption in these bands (absorption cross-section, some 4×10^{-17} cm.^{2}) may lead to pre-ionization and thus produce the E-layer in the transition region. (See Appendix, Sec. 2, Fig. 14.) The Lyman lines L_γ and L_ζ of hydrogen coincide with two of the bands. This view of the production of the E-layer, namely ionization of O_2 in the region of transition by a pre-ionization process is now the generally accepted one.

(ii) *Regions F_1 and F_2.*—Bhar in his work as mentioned above, attributed Region F_1 to the ionization of N_2 at its second ionization potential and Region F_2 to the ionization of O at its first ionization potential. The theoretical ionization distributions as obtained by him are shown in Fig. 66. It will be noticed that the shape of the resultant curve agrees with the observed distribution curve. (*Cf.* Fig. 45.) Nevertheless, spectroscopic observations on twilight flash (see Chap. X, Sec. 7) seem to throw doubt on the importance of the ionization of molecular nitrogen in the production of the ionized layers. It is maintained that the number of N_2^+ ions formed by absorption of solar radiation is but a small fraction of the total ion content of the ionosphere. Hence it is suggested that (suggestion initially advanced by Bates and Massey [127] and confirmed by later authors) Region F_1 is produced not by ionization of N_2 but by ionization of atomic oxygen. In support of this the following arguments may be advanced: We know from Chapman's theory that if the rate of electron disappearance is given by $\alpha(N_e)^2$, where α is the effective recombination coefficient, then, according to Eq. (116.1), the maximum of the layer occurs where the concentration of the active atmospheric particles has the value $\cos\chi/AH$ (A is the absorption cross-section of the ionizing radiation and H is the local scale height). The concentration of O atoms at the level of its maximum ionization is thus found to be $6{\cdot}7 \times 10^{10}$/cm.^{3} (0·9 is taken as the representative value of $\cos\chi$, $A = 4{\cdot}5 \times 10^{-18}$ cm.^{2} and $H = 30$ km.). This agrees fairly well with particle concentration as is believed to exist at the maximum of the F_1-layer. According to the above authors F_1-layer is, therefore, produced by ionization of atomic oxygen.

It is further proposed that F_1 being produced in the normal manner by photo-ionization (of O), the F_2-region is produced by a sort of bifurcation of the same according to one of the two following possible processes. The one, favoured by Mohler [126] and by Bates and Massey [127] is that the effective recombination coefficient decreases with height so rapidly that a maximum of equilibrium electron density is formed at some height above the F_1-region maximum. The single F_1-region thus appears as double with one maximum at the observed level of F_1 and another higher up (at F_2)

TABLE IV.

Characteristics of the different ionospheric regions

Region	Level of maximum ionization (km.)	Average maximum electron number density (N_e) $S = \frac{N_e \text{ in sunspot max.}}{N_e \text{ in sunspot min.}}$	Semi-thickness Y_m (km.)	Rate of ion production (I) and coefficient of recombination (α)	Number density of neutral particles	Theories about the origin—Remarks
D	60(?) (For ions; no maximum for electrons)	$N_e = 1{\cdot}5 \times 10^4/\text{cm.}^3$ Absent at night. $S = 2{\cdot}00$			$8 \times 10^{15}/\text{cm.}^3$	(i) O_2, first ionization potential (12·2 eV). Mitra, Bhar and Ghosh [129*a*]; Pfister [130]. (ii) Ionization of metals, sodium in particular. A. and E. Vassey [131]. (iii) Ionization of NO.
E	100	$N_e = 1{\cdot}5 \times 10^5/\text{cm.}^3$ (midday) $N_e = 1 \times 10^4/\text{cm.}^3$ (midnight) $S = 1{\cdot}50$	22	$I = 6 \times 10^8/\text{cm.}^3$ column $\alpha = 1 \times 10^{-8}$ cm.^{3}/sec.	$6 \times 10^{12}/\text{cm.}^3$	(i) O_2, first ionization potential (12·2 eV). Wulf and Deming [124]. (ii) O_2, second ionization potential (16·1 eV). Bhar [119]; Mitra, Bhar and Ghosh [129*a*]. (iii) $O(^1S) + O(^1S) \rightarrow O_2^+ + e$. Gauzit [132]. (iv) Pre-ionization of O_2 due to the strong absorption bands in the energy range 12·2 to 13·55 eV. Nicolet [125]. (v) High energy photon theory. Bates and Hoyle [128].

E_s	Slightly higher and overlapping Region E	Higher than E-ionization (critical frequency method of determining N_e does not hold).	Patches or thin strata of ionization embedded within the normal E		$6 \times 10^{12}/\text{cm.}^3$	Impact by meteoric particles. In high latitudes bombardment by fast solar corpuscles as produce aurorae. Appleton and Naismith [133].
F_1	200	$N_e = 2{\cdot}5 \times 10^5/\text{cm.}^3$ (midday; absent at night) $S = 1{\cdot}56$	60	$I = 1{\cdot}8 \times 10^9/\text{cm.}^2$ column $\alpha = 4 \times 10^{-9}$ cm.^{3}/sec.	$1 \times 10^{11}/\text{cm.}^3$	(i) N_2, second ionization potential. Bhar [119]; Mitra, Bhar and Ghosh [129a]; Wulf and Deming [124]. (ii) Ionization of O, at its first ionization potential (13·5 eV) giving rise to the F_1-layer at the height of the maximum ion-production, and to the F_2 layer higher up as a result of height-dependent recombination coefficient and/or tidal effect. Mohler [126]; Bates and Massey [127]; Bates [129], Martyn [134].
F_2	300	$N_e = 1{\cdot}5 \times 10^6/\text{cm.}^3$ (midday) $N_e = 2{\cdot}5 \times 10^5/\text{cm.}^3$ (midnight) $S = 4{\cdot}00$	140	$I = 1{\cdot}5 \times 10^9/\text{cm.}^2$ column $\alpha = 8 \times 10^{-11}$ cm.^{3}/sec (day) $\alpha = 3 \times 10^{-10}$ cm.^{3}/sec. (night)	$2 \times 10^{10}/\text{cm.}^3$	(i) N_2, first ionization potential (15·5 eV), Wulf and Deming [124]. (ii) O, first ionization potential (13·5 eV), Bhar [119]; Mitra, Bhar and Ghosh [129a]. (iii) Vide (ii) under F_1. (iv) N_2, first ionization potential and O second and third ionization potentials, Massey [136].

with a depression at an intermediate level. The other process advocated by Martyn, is that the bifurcation is caused by the tidal motion of the ions in the earth's magnetic field (to which the decrease of the recombination coefficient with height may also have a contributory effect). This will be explained in Sec. 14.

(iii) *Region D.*—The most widely accepted theory of the production of the D-region is that it is caused by ionization of O_2 at its first ionization potential as first suggested by Mitra, Bhar and Ghosh [129*a*].

A. and E. Vassey [131], however, maintain that photon energy down to about 7 eV is strongly absorbed by air (according to laboratory experiments) and that, as such, it is impossible for the corresponding radiation to penetrate to the required levels. Hence they suggest that the operative mechanism might be photo-ionization of atomic sodium, the threshold energy for which is only 5·1 eV. The ionization rate for this process is, however, very low and cannot explain the D-region ionization.

According to Nicolet [125] one should take account of the three following possibilities in considering D-region ionization:

(i) A normal layer due to the ionization of O_2 at its first ionization potential.

(ii) A layer more or less sporadic due to ionization of sodium.

(iii) An extra ionization during fade-out due to ionization of NO by λ 1300.

The suggestion is certainly more comprehensive than any of the others. However, in view of the uncertainties of the parameters involved, no pronouncement can be made on the comparative importance of these processes.

For convenience the origin, height and some of the other characteristics of the various ionospheric layers are collected and shown in Table IV. The sporadic E-layer (E_s), though not discussed above—because its origin is not due to photo-ionization—is included in the Table to make it complete. The origin and characteristics of E_s will be discussed in the Sec. 13.

(*e*) Variations of ionization during solar eclipse—Source of the ionizing radiation on the solar disc.

During a solar eclipse the ionizations of the E- and F_1-regions are found to decrease markedly, the minima occurring a few minutes after the obscurity [137*a*, 137*b*, 137*c*, 139, 140, 141]. For the F_2-region the effect is not so marked. While some observers report to have found a decrease in ionization (for example, references [141*a*, 139]) others have failed to detect any noticeable change [137*c*, 141*c*, 141]. According to Japanese observations, the virtual height of the F_2-layer indicated a definite decrease during the eclipse of May 9, 1948 [141]. Development of a new stratification between F_1 and F_2 was recorded in the same observations.

Ionospheric observations during a solar eclipse afford an important means of ascertaining the source of the ionizing radiation on the solar disc.

It may be recalled in this connection, that ultraviolet (black-body) radiation, as also emissions from the faculae, the chromosphere and the corona of the sun [see Appendix, Sec. 8] have all been suggested as responsible for the ionization.

During a solar eclipse, as the exposed area of the solar disc varies, the intensities of the ionizing radiations from each of the above sources also vary in manners characteristic of their distributions on the solar disc. Now, from the observed variation of ionization (say for the E-region) during an eclipse, and by making appropriate assumption regarding the value of the recombination coefficient, one can construct a curve depicting how the intensity of the ionizing radiation reaching the upper atmosphere varied in course of the eclipse. (Incidentally, this also allows one to make an estimate of the value of the recombination coefficient [176].) One may also construct curves, for each of the possible sources listed above, (knowing their distributions on the solar disc) depicting how the intensity of the ionizing radiation from each varied in course of the eclipse. A comparison of the former curve with the latter ones, then enables one to judge the most probable source of the ionizing radiation on the solar disc.

Such comparisons have been made by Waldmeier [137] with the eclipse data of July 9, 1945, from which it appears that the corona may be the source of active ionizing radiation for the E-region, though, all earlier work seemed to indicate that the ultraviolet (black-body) radiation was the ionizing agency. This is also generally corroborated by observations made in Japan during the solar eclipse of May 9, 1948 [138, 141]. Further, $h'-t$ records made on a number of fixed frequencies during the same eclipse, showed that the absorptions in the E and D regions diminished, though, the degree of the absorption and the eclipse did not go in parallel. This was ascribed probably to asymmetrical distribution of the active coronal regions on the solar disc [140].

Note :—The ionizing radiation of the corona comes from the inner corona—the parts closest to the sun. This part is believed to possess a very high temperature—of the order million degrees—and emits, beside visible radiation, radiation in the extreme ultraviolet. The corona (the part beyond the solar disc) can now be studied even when there is no solar eclipse. Intensity measurement of the corona, close to the limb of the sun is now made regularly with the green coronal line λ 5303 (due to Fe^{+13}). As the coronal structure does not undergo appreciable change in course of a week or so (at least at times of small solar activity), and, as the sun makes a quarter rotation in less than a week, it is possible to build up the structure—the intensity distribution—of the inner corona on the solar disc for any selected day from continuous observations at the limb for seven days prior and seven days after the selected day. It is to be remembered that in the observations at the limb, all elements tangential to the sun contribute to the intensity. For intensity in the front of the solar disc, the elements contained in half the corresponding column contribute. This has to be taken into account in constructing the intensity distribution diagram on the solar disc.

Mention may be made here of the suggestions of Wooley [144] and of Wooley and Allen regarding the nature of the coronal radiations which produce upper atmospheric ionization. According to these authors

F_2-region is formed by direct coronal radiation, the responsible radiation being that due to Mg^{+9}. E and F_1 regions are formed by Lyman quanta (hydrogen) from the chromosphere, but the necessary energy for this is supplied by heat conduction inwards to the chromospheric hydrogen and to the photosphere from the base of the córona which is supposed to possess a temperature of the order million degrees. According to Kiepenheuer there would be negative correlation between the ionosphere and the corona [142]. Workers in Japan, however, suggest that ionization in the F_2-region 'is maintained by two independent mechanisms relating to the solar corona in different ways. One of them is positively correlated with the green coronal line and the other negatively with it' [143].

In concluding this section we may also note the suggestion sometimes made that fast neutral corpuscles ejected from the sun may also be an ionizing source. Simple calculations show that for such particles, the time of 'corpuscular' eclipse, as also its track will be widely different from those of the optical eclipse. For a speed of 1600 km./s., for example, the optical eclipse will occur some two hours after the corpuscular eclipse and the shadow track due to the latter will lie to the east of the optical totality track at a distance exceeding 1600 km. [137*a*]. Eclipse observations, however, do not lend support to this theory.

11. DIURNAL, SEASONAL AND SUNSPOT CYCLE VARIATIONS OF THE IONIZED REGIONS

All the ionized regions of the ionosphere—D, E, F_1 and F_2—are under strong solar control. The ionizations vary markedly not only with the sun's varying zenith distance (diurnal and seasonal), but also with the spottedness of the sun in course of a solar cycle. These variations will be discussed in the present section.

(a) Diurnal and Seasonal Variations

(i) *E and F_1 regions.*—The diurnal variations of the E and F_1 regions are found to agree fairly well with the $\sqrt{\cos\chi}$ law of variation of the simple Chapman region (χ—zenith angle of the sun). In particular, the variation is found to be almost symmetrical with reference to the maximum at noon. (Fig. 67.)

The seasonal variations of the E and F_1 regions for a given hour of the day (say noon), also follow approximately the $\sqrt{\cos\chi}$ law. (For the E-region there are frequent apparent departures. These are caused by *sporadic* or *abnormal E* and will be discussed in Sec. 13.) The variation is found to be symmetrical about the summer solstice when χ attains minimum value.

This agreement with $\sqrt{\cos\chi}$ law for the E-layer may seem surprising because, this layer is produced by ionization of O_2 in the transition region where the O_2-density diminishes rapidly with height, not

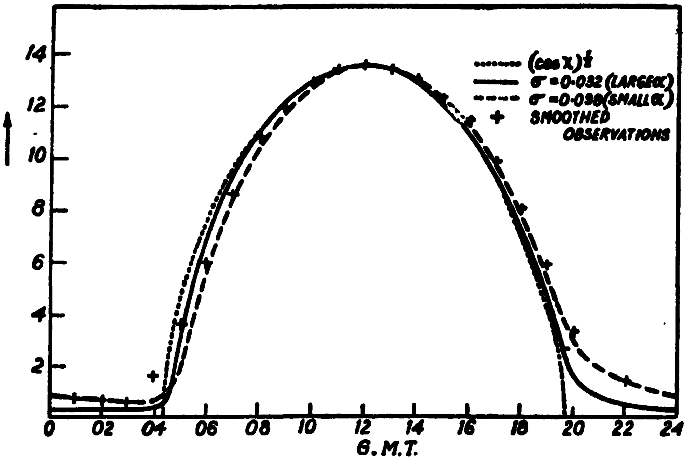


FIG. 67. Observed and calculated variations of Region *E* electron number density for summer in England. (After Best, Farmer and Ratcliffe.)

according to the simple Dalton-law (in which H remains constant with height, and which is a condition for the Chapman layer formation), but due to photo-dissociation of O_2. This anomaly can be explained if it is assumed (as has been done in Sec. 10) that the combined effects of dissociation and rising temperature in the transition region is such as to cause an exponential fall of O_2 molecule density with height. The factor in the exponent thus replaces $1/H$ and is constant.

The diurnal variation in the height $h(\chi)$ of a layer of given ionization value also follows the simple Chapman law, namely $h(\chi)$ varying linearly with log (sec χ). It is to be mentioned that with the accepted value of the recombination coefficient the E-layer ionization at night should fall to a very low value. The residual ionization density as observed however is much greater, and has not yet been adequately explained. It is sometimes suggested that bombardment by meteoric dust might replenish the night time ionization of the E-layer.

(ii) *D-region.*—Observations on D-region, which is more or less an absorbing region, are very meagre.

A layer of given electron density in the region from which long- and very-long-waves are reflected appears to undergo height variations with the zenith angle of the sun following approximately the Chapman law log (sec χ). For the regions from which higher frequencies are reflected the law ceases to hold [146].

Variation of absorption by the D-region (which is also an index of the variation of ionization density of this region) has been studied by observing the variation in the apparent reflection coefficient of the higher ionospheric regions. It is assumed that the loss in intensity of the reflected wave is caused by absorption in the double passage of the wave through the underlying D-region. If the exploring wave frequency (f) is much greater

than the collision frequency in the D-region then we have from Eq. (123), for a given value of χ,

$$(\log_e \rho)^{-\frac{1}{2}} = \text{const.}\ (f_1+f_2)$$

where ρ is the 'reflection' coefficient. According to Appleton, measurements carried out over long periods at Slough show that for signals reflected from the E, F_1 and F_2 regions the above linear relationship is closely followed [45].

Diurnal variation of absorption at vertical incidence (150 kc./s.) as measured by Benner [147] shows that the absorption coefficient (expressed by $\log \rho$) varies as $(\cos \chi)^n$, where n is 0·675 in the morning and 0·76 in the afternoon.

Long series of similar reflection coefficient measurements at normal incidence has shown that the ionization in D-region is markedly solar controlled and that there is a seasonal change of from 2 or 3 to 1 from summer to winter.

(iii) *F_2-region.*—The variations of F_2-ionization are by no means as regular as those of the E and F_1 regions. The average f^0F_2 shows com-

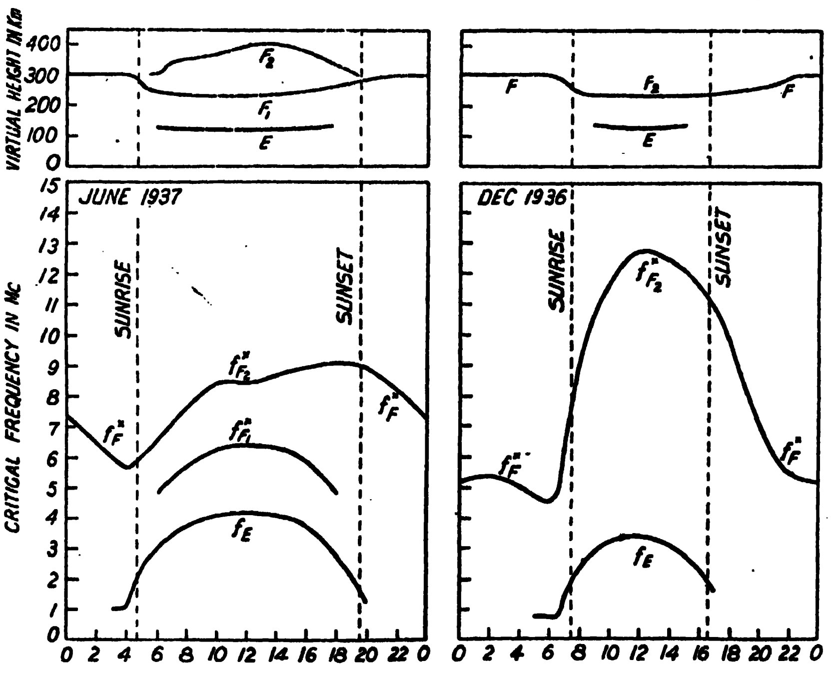


FIG. 68.—*Upper:* Average diurnal variations of the virtual heights of the E, F_1 and F_2 regions. Left: June, 1937; right: December, 1936. *Lower:* Average diurnal variations of the critical frequencies for the E, F_1 and F_2 regions. Note that the bifurcation of F into F_1 and F_2 has occurred only in the summer month. This is a characteristic of temperate and high latitude stations. Also note that the critical frequency for F_2 is lower in summer than in winter and that there is a depression about noon in summer. (The curves are for Washington. The abscissa represents Eastern Standard Time.)

plicated fluctuations not predicted by Chapman's theory [148]. The main features of the variations are noted below.

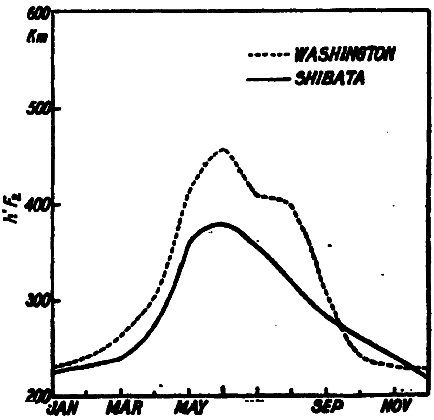
The fluctuations of ionization are much larger than those of the underlying regions—the individual observations showing a large scatter around the mean plot of f^0F_2 taken over an interval, say, a month. This scatter is probably due to sudden ionospheric storms which are more frequent in the F_2-layer [149].

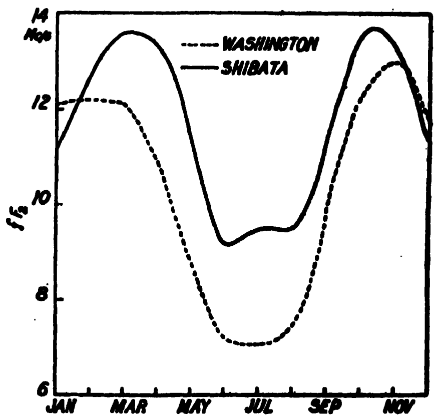
The diurnal variation is never symmetrical about noon except during winter months at mid-latitude stations. The ionization rises and falls more sharply at sunrise and sunset than the ionizations of the E and F_1 regions (Fig. 68).

In general, the diurnal variation curve has two maxima, one before and another after midday. The midday depression is greatest during summer months (Fig. 68).

The F_2-region also does not obey the simple Chapman relation regarding variation of the height of a layer of given density with the zenith angle of the sun. In winter, the height is greater at midnight than that at midday. In summer, the height is greater at midday than at midnight.

There are erratic variations of ionization during the night. It has been observed that the ionization instead of gradually diminishing during the hours of darkness (i.e., darkness in the F_2-region) often undergoes abnormal increase. It has also been found that the hour at which the increase begins, recedes towards the earlier part of the night with the approach of the winter solstice [150, 151, 152, 153].





FIGS. 69 and 70.—Illustrating *longitude effect* in the values of the average noon equivalent height (left) and average noon critical frequency (right) of the F_2-layer. Two stations situated in approximately same geographical latitude but in widely different longitudes show systematic differences in the values of these parameters. The differences are most marked when the sun's zenith distance is low. Washington—Lat. 39°N., Long. 77°W.; Shibata—Lat. 37·9°N., Long. 139·3°E. (After Appleton.)

Two stations situated in the same latitude but in widely different longitudes show systematic differences in the average penetration frequency and also in the average height of the F_2-layer (Figs. 69 and 70). This difference is particularly marked when the sun's zenith distance is low, and may sometimes amount to about 3 Mc. for f^0F_2 and about 100 kms. for $h'F_2$ in temperate latitude stations at noon. The phenomenon may be called the *longitude effect* [153*a*].

Another aspect of what is essentially the same phenomenon is as follows. The seasonal variations at two stations situated at equal *numerical* latitudes north and south of the equator and roughly at the same longitude show marked asymmetry. Thus, while the variation may have a maximum at winter solstice for some northern station, it may be so at the equinox, instead, for the corresponding southern station.

This longitude effect has been found to be associated with the earth's magnetic properties.

Geomagnetic control.—As will be presently seen, a large part of the observed anomalies of the F_2-layer can be explained as due to effects of atmospheric tides as controlled by the influence of the terrestrial magnetic field on the motions of electrons and ions. In fact, the observed F_2-region variations appear to be more under tidal-cum-geomagnetic control than under solar (ultraviolet radiation) control. (The theory of tides in the ionosphere will be discussed in Sec. 14). Here we will give an interesting instance of direct geomagnetic control of the distribution of F_2-ionization with latitude as first pointed out by Appleton [154].

It is found that for stations of widely differing longitudes, but lying not far from the geomagnetic equator the latitudinal variations of the noon value of f^0F_2 show much less scatter when plotted as a function of the magnetic dip of the station, than as a function of the geographic latitude. Figure 71 in which the f^0F_2 values for September 1948 are plotted against magnetic dip illustrates this. According to Appleton 'it appears that, for noon conditions, there is a belt of low values of f^0F_2 circling the earth and centred roughly on the magnetic equator. For stations situated within this belt it is found that these low values are associated with marked bifurcation of the F-layer into F_1 and F_2 strata. Such bifurcation is accompanied by the usual phenomena (for example, low noon value, evening concentration of ionization, slow electron disappearance after sunset, etc.) with which we are familiar under English summer conditions. In other words, the longitude effect and the geomagnetic control are exhibited not only in the noon values of f^0F_2 but also in the whole diurnal behaviour.'

It should be mentioned here that it has been rightly pointed out by Liang [155] as also by Bailey [156] that the critical frequencies should properly be plotted against geomagnetic latitude rather than against magnetic dip as in Fig. 71. This is because the dip angles as measured near the surface of the earth are often influenced by local magnetic conditions

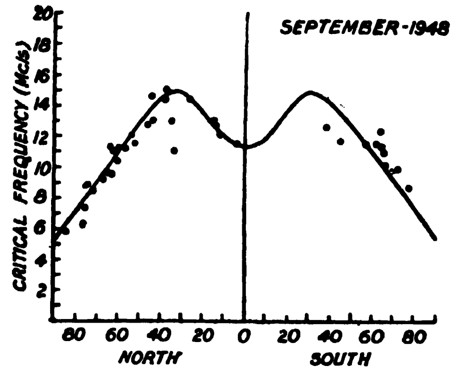


FIG. 71.—Illustrating geomagnetic control of the critical frequency (f^0) of the F_2 layer. It is found that if the average noon f^0F_2 values of stations situated in northern and southern hemispheres are plotted against the magnetic dip values of the stations (instead of the geographical latitudes) the f^0F_2 values lie more or less on a smooth curve. (After Appleton.)

and as such do not represent the geomagnetic effect in the high regions of the atmosphere.

Another feature of the geomagnetic control of F_2-layer may be mentioned. It has been reported by more than one observer [156, 157, 157*a*] that the F_2-layer situated in the vicinity of the magnetic equator is often found to be subdivided into two or more layers during daytime.

An explanation of the curious dip in the ionization values in Fig. 71 has been given by Mitra as follows [158]: In the region above the F_2-layer where the fringe of the atmosphere might be supposed to begin, the collisional frequency is very small and the electrons and ions produced by solar ultraviolet rays have very long free paths (Chapter IX, Sec. 6). They are thus free to spiral round the magnetic lines of force and, at the same time, are roughly guided along them because, when formed by photon absorption, they will in general have velocity components along the lines of force. Now, at the magnetic equator the lines of force rise highest and slope north and south. The ions and electrons formed in the high atmosphere in the belt along the magnetic equator are therefore guided north and south and, when they come down to the lower levels, contribute to the ionization density of Region F_2. The density on either side of the magnetic equator is thus increased by this 'distilling' process which operates throughout the daylight hours. It is shown that electrons and ions formed at heights 600–1200 km. above the magnetic equator will enter the earth's atmosphere at 400 km. level, in the region of magnetic dip value 19°–34°. (See Chapter IX, Sec. 6.)

That the deficiency of the ion density in the equatorial band may arise from drift of the ions to higher latitudes has also been suggested by Menzel and Bailey [57]. According to these authors, however, the motive power of the drift is daytime thermal expansion rather than solar photon absorption as proposed above by Mitra.

(b) Solar cycle variations

Observations carried out over long periods show that the ionizations of all the ionospheric regions increase or decrease as the spottedness of the sun increases or decreases in course of a solar cycle. In order to make a proper estimate of the dependence of ionization on the solar activity, it is convenient to introduce after Appleton the so-called *region character figure* as follows [159]. If N_e is the noon-ionization density (when dN_e/dt may be assumed to be zero) in an ionospheric region over a station where the zenith angle of the sun is χ, we can write

$$N_e^2/\cos\chi = q_0/\alpha$$

where q_0 is the rate of ion production when the sun is vertically overhead (noon at the equatorial region during the equinoxes). If f_0 is the ordinary ray critical frequency at the station under consideration then $N_e = Af_0^2$ where A is a known constant. We thus have

$$f_0^4/\cos\chi = q_0/\alpha A^2$$

The quantity $f_0^4/\cos\chi$ is called the region character figure for the ionospheric layer under consideration. It will be noticed that the region character figure is independent of the season, the variation in the rate of ion production due to change in the seasonal noon value of χ being taken into account by $\cos\chi$ in the denominator. The region character figure is thus a measure of the intensity of the solar ultraviolet radiation producing ionization in the region under consideration.

According to Appleton and Naismith's observations [159] the region character figure for the E-layer increased by a factor 2·2 from the year of sunspot minimum (1933-34) to the year sunspot maximum (1937-38). It appears from subsequent observations, however, that the value of the character figure is greater during years of rising activity than during years of falling activity [45]. In Fig. 72 the monthly average of the E-region

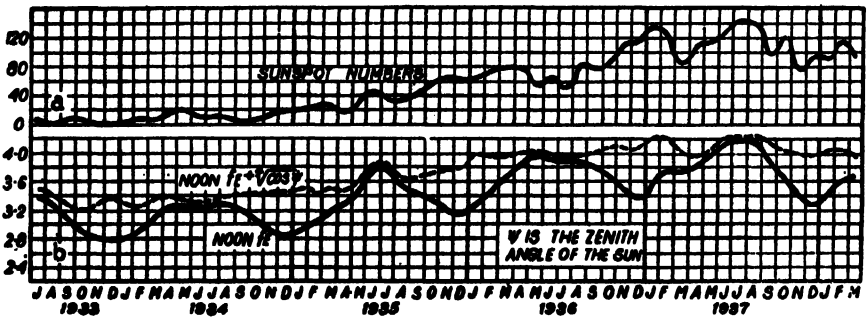


FIG. 72. Correlation between sunspot numbers and monthly average values of region character figure for Region E. (After Smith, Gilliland and Kirby.)

character figure (strictly, the fourth-root of it) for the period 1933–37 and the values of the monthly sunspot number for the same period are shown [160]. It will be seen there is general agreement between the two curves.

It will further be noticed that while the noon f^0E curve (the lower thick line curve) shows the annual seasonal variations, this has been eliminated in the upper $(f^0/\sqrt[4]{\cos\chi})$ curve. This shows that the region character figure curve provides a good basis for comparison with solar activity.

For the F_1-layer, which is well developed only during summer daylight period, the region character figure has been found, according to observations at Slough, England [45] to increase by a factor of 2·4 over the period 1934–37 (rising solar activity) and to decrease by a factor 2·6 between 1937–45 (falling solar activity).

The ionization of the D-region appears to undergo variations with the solar cycle. The variation from maximum to minimum is 2·7 : 1 in summer and 1·4 : 1 in winter.

The ionization of the F_2-region also shows marked solar cycle variation. However, the variation in this case cannot be depicted with the help of character figure because as already explained, the F_2 ionization is not related to the zenith angle of the sun in any simple manner. The winter solstice variation in course of a solar cycle is found to be much larger than the summer solstice variation.

In Fig. 73 the variations of the annual averages of the sunspot numbers

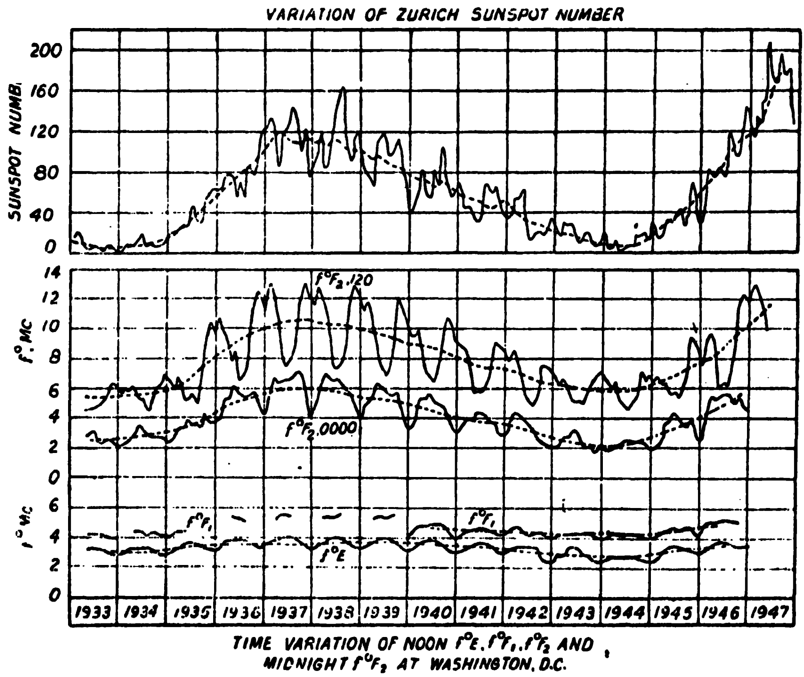


FIG. 73.—Comparisons of the variations of the noon critical frequencies for the E, F_1 and F_2 regions at Washington with the variations of the annual averages of sunspot numbers for the period 1937–47. (After Wells.)

and the variations of the noon critical frequencies for the E, F_1 and F_2 regions at Washington are compared for the period 1937–47 [162]. It will

be seen that the critical frequency for the F_2-region has increased with the sunspot numbers and that the amplitude of the variation is much greater than that for the E and F_1 regions. Thus, while f^0E has increased by a factor 1·25, f^0F_2 has increased by a factor 2·0 in the solar half-cycle under consideration.

Allen has made a close study of the average critical frequencies Rf° (relative to the critical frequency at zero sunspot number) for a number of stations for the eleven-year period 1936–47 and has correlated them with sunspot numbers and other solar features associated with an active sun [163]. It is found that Rf° for the E, F_1 and F_2 layers lag behind the sunspot numbers by 0·43, 0·66 and 0·68 parts of a month respectively. This may be interpreted as giving the order of increasing mean life of the relevant solar features.

From the above account it is clear that the critical frequencies of the various ionospheric regions are an accurate measure of solar activity—the changes in critical frequencies being associated with individual groups of sunspots. One cannot, however, expect a one-to-one correspondence between the two since there may be no direct causal relation between sunspots and the ionizing radiations, both being different manifestations of solar activity.

Waldmeier [161] has examined closely the f^0E variation with solar activity for over four years and comes to the conclusion that the source of the ionizing radiation may not lie in the sunspots but in the faculae or super-heated corona.

It is interesting to note that a 27-day variation in f^0F_2 (after correcting for lunar tidal influence) has been observed by Bartels from examination of Huancayo data [164]. The variations were found to be of the order of 6 to 10 per cent.

12. ELECTRON PRODUCTION AND ELECTRON DECAY—EFFECTIVE RECOMBINATION COEFFICIENT

In discussing Chapman's theory of ionized layer formation it was tacitly assumed that the electrons are produced only by photo-ionization and are lost only by recombination with positive ions. Under actual ionospheric conditions, however, the processes of production and loss of electrons are much more involved due to the various inelastic collision processes amongst electrons, ions and neutral particles. For example, free electrons may be lost by attachment to neutral atoms and molecules (forming negative ions, e.g., O^- and O_2^-) and may be produced by detachment from the same by collision with neutral particles or by photo-electric action. Again, the negative ions which are the source of such electrons may themselves be neutralised by electron transfer to positive ions—no free electrons being produced. The net rates of electron production and electron loss are thus determined by these and various other collisional process.

In what follows we shall first list the various inelastic collision processes amongst electrons, ions and neutral particles as are of importance in electron production and electron loss in the ionospheric regions. Brief accounts of experimental methods for determining the so-called effective recombination coefficient and the rates of electron production will then be given. Lastly, the current theories to explain the observed results will be summarized.

(a) Inelastic collisional processes

Inelastic collision processes which control the rates of production and loss of free electrons in the different ionospheric regions are the following:

(i) *Recombination.*—A positive ion may combine with an electron and get neutralized. An essential condition for recombination of an electron with a positive ion is, however, that the energy which had been spent in the ionization and which is released (at least partly) during the process of recombination must be removed from the scene of the reaction as, otherwise, the two bodies would meet just to fly off again. One of the two means of shedding this extra energy is by radiation. The probability of *radiative recombination* depends upon the probability of this energy being radiated away in the brief interval of time during which the ion and the electron are in collisional contact. The other means is by the so-called *three-body collision* process. If a third body (atom or molecule) is present in the scene of collision, then it may absorb and carry away the excess energy. The probability of recombination by this process depends upon the probability of the third body colliding with the ion and electron during the short interval of time they are in a 'state of collision'. (These remarks also obviously apply to the recombination of dissociated atoms to re-form the parent molecules. See Chap. V, Sec. 2.)

At pressures higher than 0·1 mm. the recombination proceeds mainly by the *three-body recombination* process and the coefficient of such recombination is proportional to the gas pressure [Chapter X, Sec. 7]. As such, the three-body recombination process is unimportant in the E and F regions where the pressures are of the order 10^{-3} mm. and 3×10^{-5} mm. respectively. It may, however, be of importance in the underlying D-region where the pressure is high.

There is always a finite probability for the radiative recombination process if the neutral particle produced has also a finite probability of being ionized by light absorption. According to Milne [118], the cross-section Q_a leading to ionization is related to the cross-section Q_e of the opposite process (i.e., recombination accompanied by light emission) by the formula,

$$\frac{Q_a}{Q_e}=\frac{m^2v^2c^2}{2(h\nu)^2},$$

where v is the velocity of the ejected electron and other symbols have their usual significance. (Analogous expression for the dissociation process is discussed in Chapter V dealing with the presence of atomic oxygen in the

upper atmosphere.) The following are the chief radiative recombination processes which may take place in the ionospheric regions:

$$O^+ + e \rightarrow O + h\nu \quad \text{(i)}$$
$$O_2^+ + e \rightarrow O_2 + h\nu \quad \text{(ii)}$$
$$N_2^+ + e \rightarrow N_2 + h\nu \quad \text{(iii)}$$

The recombination coefficients of all these processes are of the order 10^{-12} cm.^{3}/sec. and are listed in Table V. It is to be noted that the neutral particles produced may be in some excited state instead of in the ground state.

Besides by radiation, the extra energy on recombination, can also be accounted for by other ways. For instance, the extra energy may lift two of the outer electrons of the neutral particle produced to excited states. This is what is called *dielectronic recombination* and its probability has been discussed by Massey and Bates [165] for the case of O atom. The calculated probability is of the same order as that of radiative recombination.

For the case of positive ions of molecules there may be the so-called *dissociative recombination*, in which the energy released on recombination is spent in dissociating the molecule and, at the same time, exciting the atomic products. Thus

$$XY^+ + e \rightarrow X^* + Y^*$$

According to Bates [166] the probability of the process is extremely high and may be as much as 10^{-7} cm.^{3}/s. The dissociative recombination processes as are of importance in the ionosphere are

$$O_2^+ + e \rightarrow O^* + O^{**}$$
$$N_2^+ + e \rightarrow N^* + N^{**}.$$

No detailed calculations of these processes have been made. But it may be pointed out that for the first there is enough energy available for dissociating the O_2 molecule and, at the same time, raising one of the O atoms produced to the ^1S state and the other to the ^1D state. [Energy available is the ionization energy of O_2, namely, 12·20 eV. But the total energy used up in the process is only 11·21 eV; being the sum of the energy of dissociation (5·08 eV), the energy of excitation of ^1S (4·17 eV) and the energy of excitation of ^1D (1·96 eV)]. For the second process, the available energy falls just short of the total required, if one of the dissociated N atoms be raised to the ^2P state and the other to the ^2D state. Thus, the energy available is the ionization energy of N_2, namely, 15·58 eV. But, the total energy demanded is 15·69 eV, being the sum of the energy of dissociation (9·76 eV, Gaydon's value [165*a*]), the energy of excitation of ^2P (3·56 eV) and the energy of excitation of ^2D (2·37 eV). However, it has been pointed out by Mitra [166*b*] that the slight deficiency of 0·11 eV is made up if the N_2^+ ion be in one of the vibrational levels $v'' = 1, 2$, etc. instead of in the ground level $v'' = 0$. Alternatively, as has been suggested by Bates (private communication) the deficiency may be made up by the kinetic energy of the combining electron.

(ii) ***Mutual neutralization of positive and negative ions.***—For mutual neutralization of positive and negative ions we consider the *electron transfer* process which may have a high degree of probability. Thus

$$X^- + Y^+ \rightarrow X \text{ (excited)} + Y \text{ (excited)} + h\nu.$$

Here the electron from the negative ion X^- is transferred to the positive ion Y^+ and both particles are neutralized. The energy released is taken up partly by X and Y being raised to excited states, partly as the kinetic energies of these particles and is partly removed by radiation $h\nu$.

The coefficients of such processes are usually considered to be high (of the order 10^{-8}—10^{-7} cm.^{3}/s.) if there is energy balance, i.e. if the net amount of energy released (the ionization energy of positive ion formation less the electron affinity of the negative ion) is used up almost entirely in exciting the two neutral products. However, Bates and Massey [183] have advanced arguments to show that the maximum probability occurs not for the cases for which there is exact energy balance, but for those which are exothermic by a few volts. But, it should be mentioned that this conclusion has been questioned by Bennett and by Sayers [184].

An alternative process is

$$X^- + Y^+ \rightarrow XY \text{ (excited)} + h\nu,$$

where X^- and Y^+ combine to form a neutral molecule. This process has also high probability if $h\nu$ is small.

The possible mutual neutralization processes in the ionospheric regions are the following:—

$$O^+ + O^- \rightarrow O \text{ (excited)} + O \text{ (excited)} \quad \text{(i)}$$

$$N_2^+ + O^- \rightarrow N_2 \text{ (excited)} + O \text{ (excited)} \quad \text{(ii)}$$

$$O_2^+ + O_2^- \rightarrow O_2 \text{ (excited)} + O_2 \text{ (excited)} \quad \text{(iii)}$$

$$O_2^+ + O^- \rightarrow O_2 \text{ (excited)} + O \text{ (excited)} \quad \text{(iv)}$$

(iii) *Production of negative ions by electron attachment to neutral particles.*—Of the various constituents of the upper atmosphere only oxygen atoms and molecules form stable negative ions. (See Appendix, Sec. 2*h*.) We have, therefore, the following two processes of attachment:—

$$O + e \rightarrow O^- + h\nu,$$

$$O_2 + e \rightarrow O_2^- + h\nu.$$

Collisional experiments show that the electron affinity of O is 2·2 eV [168] and is greater than that of O_2, by about 1 eV [167]. There is some uncertainty about the value of the electron affinity of O. While Lozier gives the value as 2·2 eV, experiments by Vier and Mayer [169] have yielded the value 3·0±·1 eV. The coefficient of attachment β for O atom as calculated by Bates and Massey [167] is 1×10^{-15} cm.^{3}/sec. for incident electron of energy 0·2 eV. For oxygen molecule, whose electron affinity

is nearly half of that of oxygen atom, the attachment coefficient may be assumed to be about one order less than that of oxygen atom.

(iv) *Detachment of electrons from negative ions.*—The following reactions are possible under this head:—

$$O+O^- \rightarrow O_2+e, \qquad \text{(i)}$$

$$N_2+O^- \rightarrow N_2O+e, \qquad \text{(ii)}$$

$$O_2+O_2^- \rightarrow 2O_2+e, \qquad \text{(iii)}$$

$$O+O_2^- \rightarrow O_3+e, \qquad \text{(iv)}$$

Similar reactions with nitrogen atom, leading to the formations of NO and NO_2 may be imagined if dissociation of N_2 in the ionospheric regions is assumed. (See Chap. V, Sec. 3). There can also be the following photo-detachment processes:

$$O^-+h\nu \rightarrow O+e, \qquad \text{(v)}$$

$$O_2^-+h\nu \rightarrow O_2+e. \qquad \text{(vi)}$$

Processes (i) to (iv) represent detachment by collision; these processes are always operative. Processes (v) and (vi) represent photo-detachment and they may play an important rôle during daytime. An estimate may be made of the cross-section for the detachment process represented by (i) from the following consideration. When a beam of electrons is fired through a gas such as O_2 the reaction is $O_2+e \rightarrow O+O^-$ which is the reverse of (i). The cross-section of O_2 for this reaction has been estimated to be not greater than 5×10^{-19} cm.^{2} Application of the principle of detailed balancing shows that the cross-section for the opposite process (i) will be small, only of the order of 10^{-22}—10^{-21} cm.^{2} [170]. This is an inferior limit, the actual value may be greater by a factor of about 10. This is because in the electron firing process the oxygen molecules fired at are in the ground state, but in reaction (i) the association of O and O^- may produce also vibrationally excited molecules. The value of the detachment cross-section corresponding to process (i) may, therefore, be assumed to lie between 10^{-21} to 10^{-20} cm.^{2} The coefficients of the reactions (ii), (iii) and (iv) have not been calculated. It may, however, be assumed that the coefficient of reaction (ii) is of the same order as that of reaction (i). For reactions (iii) and (iv) the values of the coefficients may reasonably be taken higher than the detachment coefficient of electrons from O^- because the electrons are more loosely bound to O_2^-.

In the photo-detachment processes (v) and (vi), negative ions eject electrons by absorption of light quanta having energy greater than the electron affinity. The rate of production of electrons from negative ions (say I) by light absorption may be evaluated as follows [171]. We have

$$= N_- \int_{\nu_1}^{\infty} Q_a(\nu) n_\nu d\nu,$$

where n_ν—the number of light quanta of frequency ν passing through unit area per unit time,
N_-—the number of negative ions per cm.^{3},
ν_0—light frequency corresponding to the energy of attachment,
$Q_a(\nu)$—cross-section for absorption of light of frequency ν.

Using Milne's relation as given above and assuming that for the low velocity electrons as are present in the ionosphere Q_a varies as $1/v$ [172], we obtain [see Appendix, Sec. 6]

$$I = N_- \gamma_1 \beta,$$

where γ_1 is a constant given by

$$\gamma_1 = \psi^2 \frac{(2\pi m)^{\frac{1}{2}}}{4h^3} k^{\frac{1}{2}} T_e^{\frac{1}{2}} e^{-h\nu_0/kT_e}.$$

For O atoms, taking electron affinity $= 2{\cdot}2$ eV, γ_1 is equal to 1×10^{14}; for O_2 molecule, taking electron affinity $= 1{\cdot}0$ eV, γ_1 is equal to 9×10^{14}.

The coefficients of the different types of reactions which are of importance in the upper atmosphere are collected in Table V.

(*b*) Estimation of effective coefficient of recombination and rate of electron production from ionospheric data

The effective coefficient of recombination (α) and the rate of electron production (q) are interrelated. The magnitudes of these quantities may be estimated from a study of the diurnal ionization variation curves round noon when the ionization is nearly maximum and round sunrise when it is a minimum.

Consider points a and b in Fig. 74 immediately before and after sunrise

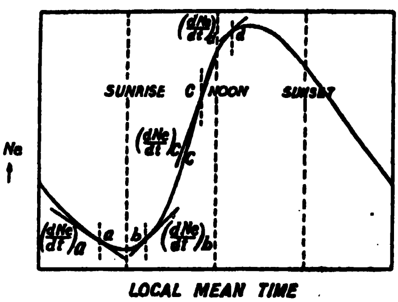


FIG. 74.—Reference points (a, b, c, d) in the diurnal ion number density variation curve for calculation of the rates of ion production.

when the electron concentrations (N_e) are the same and the values of α may also be assumed to be the same. We, therefore, put

$$\alpha_a = \alpha_b \quad \ldots \quad (137)$$

when $$(N_e)_a = (N_e)_b. \quad \ldots \quad (138)$$

We, then, have

$$q_b = \left|\frac{dN_e}{dt}\right|_b - \left|\frac{dN_e}{dt}\right|_a. \quad \ldots \quad (139)$$

TABLE V

Coefficients of the different types of reactions important in ionospheric regions

Type of reaction	Reaction	Coefficient of the reaction	REMARKS
Radiative recombination ..	$O^{+}+e \rightarrow O^{*}+h\nu$	$\alpha_e = 1{\cdot}5 \times 10^{-12}$ cm.^{3}/sec. at temp. 1000°K.	Calculated by Bates and others [166a] taking into account contributions from ground and excited states.
	$\left.\begin{array}{l} O_2^{+}+e \rightarrow O_2^{*}+h\nu \\ N^{+}+e \rightarrow N^{*}+h\nu \\ N_2^{+}+e \rightarrow N_2^{*}+h\nu \end{array}\right\}$	$\alpha_e = 10^{-12}$ cm.^{3}/sec.	No calculations have been made of these reactions but they may be assumed of the same order as above [166a].
Dielectronic recombination	$O^{+}+e \rightarrow O^{**}$	Same order as that of radiative recombination.	In this type of reaction, energy released on recombination lifts two of the outer electrons of the neutral atom produced to excited states. The probability of the process for O atom is discussed by Massey and Bates [165].
Recombination leading to dissociation and excitation.	$O_2^{+}+e \rightarrow O^{*}+O^{**}$ $N_2^{+}+e \rightarrow N^{*}+N^{**}$	10^{-8}—10^{-7} cm.^{3}/sec.	Detailed calculations have not been made. But according to Bates [166] the probability is very high. This process may be the one producing atomic nitrogen in the upper atmosphere [166b].
Mutual neutralization ..	$N_2^{+}+O^{-} \rightarrow N_2^{*}+O^{*}$	$\alpha_i = 10^{-7}$ cm.^{3}/sec.	This reaction has been considered in Chapter X, Sec. 6 (*b*) in connection with the excitation processes of the night sky spectrum [173]. The resonance condition is almost exact and the value of its coefficient, from a comparison of the preceding reaction, may be assumed to lie near to 10^{-7} cm.^{3}/sec. (See, however, text).

	$O_2^+ + O_2^- \rightarrow O_2^* + O_2^*$ $O_2^+ + O^- \rightarrow O_2^* + O^*$	$\alpha_i = 10^{-7}$ cm.^{3}/sec.	Similar to the preceding reaction; the coefficient may be assumed to be of the same order.
Electron attachment ..	$O + e \rightarrow O^- + h$	$\beta = 1 \times 10^{-15}$ cm.^{3}/sec. for electrons of energy 0·2 eV.	Calculated by Bates and Massey [167].
	$O_2 + e \rightarrow O_2^- + h\nu$	$\beta = 1 \times 10^{-16}$ cm.^{3}/sec.	The electron affinity of O_2 (1 eV) is nearly half of that of O (2·2 eV). The attachment coefficient may be assumed to be one order less than that of O.
Electron detachment ..	$O^- + O \rightarrow O_2 + e$	$\gamma = 10^{-16} - 10^{-15}$ cm.^{3}/sec.	Calculated from the reverse process [170].
	$O^- + N_2 \rightarrow N_2O + e$	$\gamma = 10^{-16} - 10^{-15}$ cm.^{3}/sec.	The coefficients of these processes have not been calculated. It is, however, assumed that they are of the same order as the preceding process. For the case of detachment of electrons from O_2^- the value may reasonably be assumed to be higher than that of O− because the electron here is much more loosely bound.
	$O_2^- + O_2 \rightarrow 2O_2 + e$ $O_2^- + O \rightarrow O_3 + e$	$\gamma = 10^{-15} - 10^{-14}$ cm.^{3}/sec.	
Photo-detachment ..	$O^- + h\nu \rightarrow O + e$	$\gamma_1 = 1 \times 10^{14}$, $\beta = 1 \times 10^{-15}$ cm.^{3}/sec.	The rate of photo-detachment is $\gamma_1 \beta N_-$ where N_- is the density of negative ions.
	$O_2^- + h\nu \rightarrow O_2 + e$	$\gamma_1 = 9 \times 10^{14}$, $\beta = 1 \times 10^{-16}$ cm.^{3}/sec.	Calculated by Tukada [171].
Photo-ionization ..	$O + h\nu \rightarrow O^+ + e$	Cross-section = $2{\cdot}8 \times 10^{-17}$ cm.^{2} ,, ,, = $2{\cdot}6 \times 10^{-18}$ cm.^{2} (λ 1019) ,, ,, = $3{\cdot}7 \times 10^{-18}$ to $3{\cdot}3 \times 10^{-17}$ cm.^{2}	Calculated by Saha and Rai [122]. Calculated by Bates and Seaton [174]. Calculated by Tukada [171].
	$N_2 + h\nu \rightarrow N_2^+ + e$	,, ,, = 1×10^{-17} cm.^{2} (at λ 795)	Approximate value [183].
	$O_2 + h\nu \rightarrow O_2^+ + e$	,, ,, = 1×10^{-20} cm.^{2} (at λ 1019) = 10^{-16} to 10^{-17} cm.^{2} (at λ = < 744).	Approximate values [183].
	$N + h\nu \rightarrow N^+ + e$	,, ,, = 9×10^{-18} cm.^{2} (λ 847).	Calculated by Bates and Seaton [174].

Next consider points c and d equally spaced on either side of noon. We can assume that q and α are the same at these points. Then

$$\left(\frac{dN_e}{dt}\right)_c = q_c - \alpha_c (N_e)_c^2 \quad \ldots \quad \ldots \quad \ldots (140)$$

and,

$$\left(\frac{dN_e}{dt}\right)_d = q_d - \alpha_d (N_e)_d^2 . \quad \ldots \quad \ldots \quad \ldots (141)$$

Hence,

$$q_c = q_d = \frac{\left[(N_e)_d^2 \left(\frac{dN_e}{dt}\right)_c - (N_e)_c^2 \left(\frac{dN_e}{dt}\right)_d\right]}{(N_e)_d^2 - (N_e)_c^2} \quad \ldots \quad \ldots (142)$$

The values of q and α for all hours of the day may now be calculated with the help of Eqs. (139) and (140).

The rapid fall and rise of E ionization as observed during a solar eclipse affords another means of computing the effective recombination coefficient in this region [176].

At the instant when f is the fraction of the exposed disc of the sun and χ is the zenith angle of the sun, the rate of electron decay is obviously given by

$$\frac{dN_e}{dt} = -\alpha N_e^2 + f q_0 \cos \chi$$

where α is the recombination coefficient and q_0 is the maximum rate of electron production. Hence, if the variations of $\frac{dN_e}{dt}$, f, and χ with time in course of the eclipse are known, one can draw a family of curves showing the variation of N_e with time for different assumed values of α. From such a family of curves the one that fits best with the observed variation may be picked out and the α corresponding to this curve may be taken as representing the most probable value of α for the E region.

Note :—q_0 is obtained from observations on control days and the variation of f may be calculated from the known characteristics of the eclipse. The method of obtaining the variation of $\frac{dN_e}{dt}$ is rather laborious but may be obtained step by step by considering the ionization variation over short intervals of time [176, 177].

A third method of estimating α for Region E consists in drawing a number of diurnal ionization variation curves like those in Fig. 67 for different values of α and selecting the one which fits best with the experimentally observed one. The value of α corresponding to this curve of best fit represents approximately the recombination coefficient of Region E.

Measurements on the rate of electron production q show that contrary to expectation its values for both E and F_2 regions are generally high after sunrise and before sunset, and low round 10 hours and 14 hours. This is specially true for stations situated near geomagnetic equator for example,

Huancayo [178] and Calcutta [55] in summer and equinox. The average values of q as measured under ionospheric conditions at Calcutta are as follows: For Region F_2, q varies between 150/cm.^{3}/s. and 100/cm.^{3}/s. In summer during high solar activity q attains a maximum value of about 400/cm.^{3}/s. For Region E the value varies between 50/cm.^{3}/s. and 30/cm.^{3}/s.

It is to be noted that the determinations of q and α as discussed above do not take into account the very important effect of vertical ion drifts in ionosphere caused by atmospheric tides. This is specially important for the F_2-region where the tidal effects cannot be ignored. A. P. Mitra [179] has determined its value for Calcutta taking into account ion drifts due to tidal motions. The value of α as determined by him is found to be of the order 10^{-11}/cm.^{3}/s. The summer value was found to be less than the winter value.

We may summarize here the values of the recombination coefficients for the main ionospheric layers as obtained by different workers [55, 180, 181, 182],

$$\begin{aligned} \alpha E &\fallingdotseq 1\times 10^{-8} \text{ cm.}^3\text{/s.} \\ \alpha F_1 &\fallingdotseq 4\times 10^{-9} \text{ cm.}^3\text{/s.} \\ \alpha F_2 &\fallingdotseq 8\times 10^{-11} \text{ cm.}^3\text{/s. (day)} \\ &\fallingdotseq 3\times 10^{-10} \text{ cm.}^3\text{/s. (night).} \end{aligned}$$

The important point to note here is that the computed values of the recombination coefficient are much greater than the probable theoretical value of 10^{-12} cm.^{3}/s. for radiative recombination. Further, the value for the lower regions E and F_1 is higher than that for the higher region F_2. It is to be noted that for these lower regions the value of α is fairly constant. But for the F_2-region the recombination coefficient decreases with height. In the next section we shall discuss the probable causes of the high value of the recombination coefficient particularly in the lower regions and also the cause of its variation with height.

(c) Equilibrium in the ionized layers

The explanation of the high value of the recombination coefficient that was first given and that was widely accepted was one in which negative ions of atomic oxygen played an important rôle (Bates & Massey [183], Sayers [184]). However, it now appears that the process of dissociative recombination can also explain the observed high value at least for the E-region if O_2 molecule be the active constituent in this region. In what follows we shall first give in some detail the theory of effective recombination as it will be helpful to understand the interrelations between the various electron production and electron decay processes in the ionospheric regions.

(i) *Effective Recombination.*

Consider the processes by which free electrons are lost and produced. They are lost (i) by recombination with positive ions, and (ii) by attachment

to neutral particles like O atoms and O_2 molecules which have high electron affinity. They are produced, (i) by detachment of electrons from negative ions by collisions with neutral particles, (ii) by photo-ionization of neutral particles, and (iii) by photo-detachment of electrons from negative ions. (The two last-named processes are, of course, operative only during daytime.) Again, negative ions are produced by attachment of electrons to neutral particles and are lost (i) by detachment by collisions with neutral atoms and molecules, (ii) by photo-detachment, and (iii) by mutual neutralization when a negative ion meets a positive ion. The rates of production and loss of negative ions thus affect the net rates of production and loss of free electrons.

For the time variation of the negative ion density at daytime we may put,

$$\frac{dN_-}{dt} = -\alpha_i N_- N_+ + \beta N_e n - \gamma N_- n - \gamma_1 \beta N_-, \qquad .. \quad (143)$$

where α_i—coefficient of mutual neutralization of positive and negative ions.

β—coefficient of attachment of electrons to neutral atoms and molecules.

γ—coefficient of detachment of electrons from negative ions by collisions.

n, N_e, N_-, N_+—number densities of neutral particles, electrons, negative ions and positive ions respectively.

$\gamma_1 \beta N_-$—rate of loss of negative ions by photo-detachment of electrons. [See Appendix, Sec. 6.]

Let us for the moment assume that the term $\alpha_i N_- N_+$, the rate of disappearance of negative ions by mutual neutralization, can be neglected compared to the attachment and detachment rates $\beta N_e n$ and $\gamma N_- n$. The assumption may be justified by the fact that the last-named terms involve n, the number density of neutral particles which is very large compared to N_- or N_+. Also, $\alpha_i N_- N_+$ may be neglected compared to $\gamma_1 \beta N_-$ because γ_1 has a very high value. Since the attachment and detachment rates are great, a dynamical equilibrium is quickly established for these rates and in the equilibrium condition we have,

$$\beta N_e n = \gamma N_- n + \gamma_1 \beta N_-, \qquad .. \quad .. \quad .. \quad (144)$$

or,

$$\frac{N_-}{N_e} = \frac{\beta n}{\gamma n + \gamma_1 \beta} = \lambda \text{ (say)}. \qquad .. \quad .. \quad (145)$$

Now, for the time variation of electron density at daytime we have,

$$\frac{dN_e}{dt} = q - \alpha_e N_e N_+ - \beta N_e n + \gamma N_- n + \gamma_1 \beta N_-, \qquad .. \quad (146)$$

where, α_e—coefficient of recombination of electrons with positive ions.

q—rate of electron (or positive ion) production per cm.^{3} due to photo-ionization.

Combining Eqs. (143) and (144) and assuming that the ionospheric regions are electrically neutral, so that,

$$N_+ = N_e + N_- = (1+\lambda)N_e,$$

and using relation (145) we have,

$$\frac{dN_e}{dt} = \frac{q}{1+\lambda} - (\alpha_e + \lambda\alpha_i)N_e^2 - \frac{N_e}{1+\lambda}\frac{d\lambda}{dt}.$$

Since γ_1 depends on $\cos\chi$, where χ is the zenith distance of the sun, the variation of λ with time will be small except during sunrise and sunset. Therefore,

$$\frac{dN_e}{dt} = \frac{q}{1+\lambda} - (\alpha_e + \lambda\alpha_i)\, N_e^2,$$

$$\text{where} \quad \lambda = \frac{\beta n}{\gamma n + \gamma_1\beta} \text{ (daytime).} \qquad \ldots \quad \ldots \quad (147)$$

At night $q = 0$, $\gamma_1 = 0$ so that,

$$\frac{dN_e}{dt} = -(\alpha_e + \lambda\alpha_i)\, N_e^2, \qquad \ldots \quad \ldots \quad \ldots \quad (148)$$

$$\text{where} \quad \lambda = \frac{\beta}{\gamma} \text{ (night time).} \qquad \ldots \quad \ldots \quad (149)$$

We thus obtain the law of electron decay by recombination, namely, the rate of decay is proportional to the square of electron number density. The quantity $(\alpha_e + \lambda\alpha_i)$ is therefore called *effective recombination coefficient*. The recombination rate is thus controlled by the negative ion—electron number density ratio λ and the coefficient of mutual neutralization α_i, rather than by the simple recombination coefficient α_e. Now, the coefficient of mutual neutralization is of the order 10^{-7} to 10^{-8} cm.^{3}/sec. and the value of λ varies according to the region and time of observation between unity and 10^{-2}. Hence the effective recombination coefficient varies roughly between 10^{-8} and 10^{-10} cm.^{3}/sec.

The expression $(\alpha_e + \lambda\alpha_i)$ for the effective recombination coefficient was first deduced by Appleton and Sayers [184] and also independently by Massey and his collaborators [166*a*, 183].

Ghosh has carefully considered the equilibrium condition of the different processes in the different ionospheric regions and the variations in the value of effective recombination coefficient resulting therefrom [185]. It has been pointed out by him that one is not justified in neglecting the term $\alpha_i N_- N_+$ in Region *F* where the number density of neutral particles is low but that of electrons high. Taking this into account he has deduced the expression for λ for Region *F* both for the day and for the night time conditions. According to him, the correct expression for λ for Region *F* for daytime is

$$\lambda = \frac{\beta n}{(\gamma n + \gamma_1\beta + \alpha_i N_e)} \qquad \ldots \quad \ldots \quad \ldots \quad (150)$$

or, since γn is many orders smaller than $\gamma_1\beta$ and $\alpha_i N_e$,

$$\lambda = \frac{\beta n}{(\gamma_1\beta+\alpha_i N_e)}. \quad \ldots \quad \ldots \quad \ldots \quad (151)$$

For night time

$$\lambda = \beta n/\alpha_i N_e. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (152)$$

Substituting the value of λ in the night time expression of dN_e/dt, it is easily seen that the rate of decay of electron number is proportional to N_e rather than to N_e^2. This has been tested by Ghosh by comparing the decay of electron number density of Region F at night as observed at Huancayo and Watheroo with those calculated according to the values of λ given by Eq. (152). He found that the calculated values agree fairly well with the observed values.

A more general expression for λ, applicable to all the ionospheric regions, as deduced by Bates and Massey [183] is

$$\lambda = \frac{\beta n - \dfrac{1}{1+\lambda}\dfrac{d\lambda}{dt}}{\gamma_1\beta+\gamma n+q+N_e(\alpha_i-\alpha_e)}. \quad \ldots \quad \ldots \quad (153)$$

For E-layer the equation simplifies to

$$\lambda = \frac{\beta n}{\gamma_1\beta+\gamma n}, \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (154)$$

and for F_2-layer to

$$\lambda = \frac{\beta n}{\gamma_1\beta+\alpha_i N_e}. \quad \ldots \quad \ldots \quad \ldots \quad (155)$$

The expression for Region F_2 is the same as that given by Ghosh. The expression for Region E, however, is different from that previously given.

It will be noticed from the equations that the values of recombination coefficient for both the Regions E and F_2, vary with height [55, 178, 186]. For Region E this variation is due to the decrease of n with height, while for Region F_2, it is due to the variations in n as well as in N_e. The variation of α with height in F_2-layer has been attributed by Baral and A. P. Mitra [55] to the variations in both temperature and pressure with height. (It may be mentioned that according to many authors the variation is due to pressure only, while according to Seaton [178] it is due to variation in temperature only.)

It may also be noticed that the law of recombination, namely, that the decay of electron density is proportional to the square of electron density, holds for Region E both at day and night and for Region F_1 (daytime). For Region F_2 during daytime, the law may not hold in the extreme case, when the effect of photo detachment is small compared to that of the effect of mutual neutralization. In such case, the decay of electron density of Region F_2 at daytime may vary linearly as N_e. A consequence of this is that the region of maximum electron production may not coincide with that of maximum electron number density.

(ii) *Dissociative Recombination.*

As mentioned in the beginning of this section the process of dissociative recombination for the O_2^+ ions can explain the high value of the recombination coefficient. According to Bates [166] the dissociative recombination process such as $O_2^+ + e \rightarrow O^* + O^*$ has a very high probability. Here the energy of ionization is spent partly in dissociating the neutralized O_2 molecule and partly in exciting the O atoms produced. Hence if O_2 molecules be the active substance we have no need of assuming any complicated effective recombination process as discussed above. We may simply write

$$\frac{dN_e}{dt} = q - \alpha(d) N_e n(\mathrm{O}),$$

where $\alpha(d)$ is the dissociative recombination coefficient and is of the order 1×10^{-8} cm.^{3}/s. For the higher regions F_1 and F_2, however, it has not been possible to suggest any such simple process.

(*d*) Laboratory method of determining recombination coefficient

It may be interesting to note that microwave techniques have been applied for developing laboratory methods for estimating rates of decay of ionization from change in resonant frequency of a microwave cavity enclosing the electrons [187, 188]. As in most of the gases used, negative ion formation is negligible, the decay of ionization is to be attributed largely to electron recombination. An important technical problem in connection with such laboratory determination is the elimination of the wall effect. For the microwave technique this is accomplished by keeping the gas pressure sufficiently high (10^{-1} mm. of mercury, remembering that at least for monatomic gases α is independent of pressure). The measured rates of recombination are found to be several orders higher than those previously reported, ranging from 10^{-8} cm.^{3}/s. to 10^{-6} cm.^{3}/s.

Another interesting method of controlling ion-diffusion to the walls which is being developed by the Research Laboratory of Physical Electronics, Tufts College, Medford, Massachussetts, may be mentioned. A discharge tube in the form of a toroidal ring is used and a longitudinal magnetic field is applied through the tube everywhere perpendicular to the cross-section to control ion-diffusion to the walls [189].

(*e*) Existence of negative ions—'Sunrise effect' (Ozone shadow)

In the preceding section we have referred to the formation of negative ions (O^-, O_2^-) in the ionospheric regions. Observations on the 'sunrise effect', i.e., on the hour at which the decay of ionization in an ionospheric layer (which proceeds during the dark hours of night) is first checked, in relation to the hour at which the solar rays tangent to the ground first strikes the layer in the small hours of the morning, afford a means of detecting the presence of such ions. It is clear that the grazing solar rays are deprived not only of the extreme ultraviolet wavelengths (by absorptions

by N_2, O_2 and O) in the high atmosphere, but also of the near ultra-violet (by ozone in the middle atmosphere) and most of the visible radiation except red (by haze and fog near the ground). If, therefore, the ionization is found to increase when the ground-grazing rays strike the layer, the electrons produced cannot be due to ionization as no atmospheric constituent has an ionization potential corresponding to red rays. The increase may, however, be due to electron detachment from the negative ions which are known to have quite low detachment potentials (see Appendix, Sec. 2*h*). Observations on the sunrise effect are thus useful in checking the possible presence of negative ions. We give below results of some observations on the 'sunrise effect' in different parts of the world.

From analysis of records made at Calcutta for about 14 months (1938-39) it was concluded that *E*-region ionization begins to increase when the solar rays strike it, not tangent to the ground, but to the ozonosphere at a height of 35 km. [190]. Further, it was found that the seasonal variation of the hour of sunrise effect, in relation to the hour of ground sunrise, followed closely the seasonal variation of the ozone content of the atmosphere. (Fig. 75.) Hence it was concluded that the observed

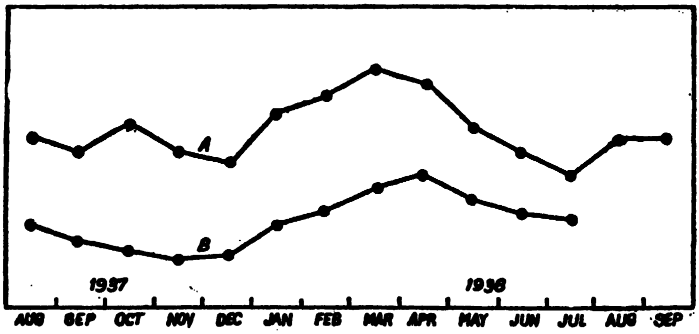


FIG. 75. Correlation between the seasonal variations of atmospheric ozone content (curve *A*) and the hour of early morning increase of Region *E*-ionization (curve *B*).

sunrise effect is due to electron detachment (possibly from O^-) and that the active wavelength is greater than 2900 A which is the ozone-cut off wave-length (see Fig. 1, Chap. IV).

Penndorf [191] has also discussed how the upper boundary of the ozone layer (responsible for delaying the sunrise effect in the *E*-layer) can be determined from the known values of (1) height of the *E*-layer, (2) time of sunrise for visible light in the *E*-layer, and (3) the time at which the ionization begins to increase. The accuracy of determination is found to be as good as $\pm$ 1 km. and the observed delay time agrees best with the calculated values if the height of the boundary of the active ozone layer is taken to be 50 km. instead of 35 km. as assumed above.

The results obtained at Calcutta, referred to above, were for the region near the maximum ionization of the *E*-layer. These may be compared with the sunrise effect for a region below the maximum as observed with 325 kc./s.

wave at Stanford, California (U.S.A.) [192]. It was found that the intensity of reflection tended to increase at about the time (which was up to an hour before ground sunrise) when the solar rays tangent to the ground struck the region. This increase has been attributed to ionization by electron detachment from negative ions. Further, the height of the lowest observed E-region reflection (on 100 kc./s.) was found to drop appreciably (the magnitude averaging 12 km. in winter and 6 km. in summer) at about the time of ground sunrise. This change is attributed to effect of ionization by solar ultraviolet radiation.

Some of the results of sunrise effect as observed for the case of the F-layer are as follows. From observations made at Calcutta [190] it was found that the pre-sunrise minimum occurred in all seasons after F-region sunrise but before ground sunrise. The hour of minimum was farthest removed from the hour of F-region sunrise (nearest to ground sunrise) in summer, but approached the same in winter. According to observations at Slough the minimum occurred prior to ground sunrise in all seasons [193]. At Washington, on the other hand, the pre-sunrise minimum was found to occur after ground sunrise during the period of observation May-June 1949 [194]. According to observations at Stanford [192], on frequencies 325 kc./s. and above, there is an abrupt drop in the virtual height of reflection (as shown by h'–t recordings) when the solar rays are incident on the region tangent to the earth's atmosphere at heights greater than roughly 200 km. above sea-level. From this it was concluded that the early morning increase in F_2 is not caused by solar radiation in the visible region.

13. IRREGULARITIES AND ABNORMALITIES IN THE IONOSPHERE

(a) Introduction

The diurnal, seasonal and solar cycle variations of ionization discussed in Section 11 may be regarded as regular and normal ionospheric variations, being dependent on regular and predictable variations of the ionizing solar radiation. Besides these regular variations there are irregular and abnormal variations in ionization which reveal themselves by sudden increase in absorption of downcoming waves, abnormal increase in the critical frequency, and/or by scattering of radio waves. The origin of some of these abnormalities is still ill-understood. Some have been traced to effects of solar flares, meteoric impacts, solar corpuscle bombardment (as produce auroras) and, probably, run-away electrons from thunder-clouds and electric field of the same. It is to be remarked that the irregularities and abnormalities may be transitory or of a semi-permanent character. For example, the ionization produced along a meteor track may be dispersed in a fraction of a second; the ion clouds which produce abnormal or sporadic E in temperate latitudes, as also the effects of extra ionization produced during a solar flare may last for an hour or so; and, the abnormal effects produced during magnetic storms may continue for days.

In regard to the last named effect, it may be mentioned that near the auroral zone where the ionizing solar corpuscles concentrate the abnormality is the normal feature of the ionosphere. In what follows we will discuss in some detail these various irregularities, their effects and their probable origins.

(*b*) Sporadic *E*

A very persistent abnormality of the ionosphere is what is known as abnormal or sporadic *E* denoted by E_s. It has been known since the early days of pulse technique that the *E*-layer critical frequency undergoes sudden and abnormal increases. The increases may be many times the normal *E* critical frequency and may sometimes be so high as even to screen the *F*-echoes. The virtual height of these abnormal echoes varies between 90–120 km.

The sporadic *E*'s occurring in different latitudes may be of different types. However, there are two types, the long duration or the 'intense' type and the transient or the 'burst' type, which are found to occur in all latitudes. Further, as the variation of the penetration frequency fE_s is dependent upon the power of the transmitter, the sensitivity of the receiver and the directional properties of the transmitting and the receiving aerials, it is not possible for ionospheric observatories to report median or average values of fE_s. Instead, the practice is to report the percentage of times fE_s is above a certain value (say 3, 5 or 7 Mc./s.).

Various considerations show that the increase in the *E*-layer penetration frequency due to E_s is caused by intrusions of localized regions of increased ionization into the body of the *E*-layer and not due to an increase in the ionization of the *E*-layer proper. In equatorial regions E_s-ionization is weak but exists throughout the daylight hours and fE_s varies smoothly with a maximum at noon [195]. In temperate and subtropical latitudes, E_s occurs most frequently in the early morning and in the evening and is, therefore, sometimes called nocturnal-*E*. It has also got a seasonal variation, both in equatorial and in temperate latitudes, being most numerous in local summer [195, 196]. E_s occurrence frequency does not seem to depend on the solar cycle. In high latitudes near the auroral zone, E_s occurs with great frequency, but its origin is believed to be different from that in the lower latitudes. The aggregate of ion clouds which produce the sporadic *E* can also scatter radio waves (*vide infra*). It may be mentioned here that observations by pulse technique with low frequency (100, 325, 350 kc./s.) show that the night-time *E*-region consists of several strata of ionization [192]. These, both transmit and reflect the low frequency waves employed, showing that they are very thin. (Measurement showed that the ionization gradient corresponds to a scale height of only 0·7 km.) A close correlation is found to exist between the highest virtual height recorded with 325 kc./s., and the minimum virtual height of E_s as recorded with the usual automatic ionospheric equipment (1·3–18·0 Mc./s.)

It is thus possible that the strata observed with the low frequencies are also sporadic-E ion clouds which, on account of their extreme thinness, though of high electronic density, transmit the low frequency waves employed. (See Sec. 3.)

The origin of E_s is not yet completely known. Meteor ionization, thunderstorm effect and bombardment by solar corpuscles (specially near auroral zones) have all been suggested as contributing to the production of E_s. We discuss below in some detail the probable effects of these agencies on upper atmospheric ionization.

(i) *Meteor ionization.*—Of the three possible sources listed above, meteor ionization has been the most favoured one, at least for low and middle latitudes. In Chap. IIIA, Sec. 5, it has been explained how meteors in rushing into the atmosphere produce trails of ionization which may be detected as transient echoes with the 'radar' technique on very high frequencies. It has been suggested that meteor ionization also produces the ion-clouds which cause the E_s echoes. Many observers have recorded temporary increase of E-region ionization density to very high values during meteoric showers [197, 198, 199, 200]. Closer study of this phenomenon has been made by Appleton and Naismith [196] in England. These authors compared the incidence of E_s echoes (as recorded in automatic multifrequency ionosphere apparatus) with that of transient meteor echoes by the 'radar' technique (using vertically directed beam on 27 Mc./s.) during the same period. In particular, they compared the occurrence frequencies of the two types of echoes during the Giacobinid meteor shower of 9-10 October, 1946, and found that during the shower period the frequency of the transient echoes, as also that of the E_s echoes increased enormously. It has also shown by Pierce [201] that during this shower reflecting ion clouds were produced, which lasted for some hours on a frequency of 3·5 Mc./s. Appleton and Naismith [196] have also shown that the seasonal variation of transient meteor trail echoes follows a trend very similar to that of the occurrence of E_s. However, these findings are questioned as a result of statistical analysis of observational data on meteor trail echoes as recorded at the Central Radio Propagation Laboratory, National Bureau of Standards, U.S.A. [202]. The echoes as observed on 27·2 Mc./s. have been compared with the incidence of long duration E_s echoes over the period—November, 1948 to April 30, 1949. From the diurnal variation curves of the two phenomena it is concluded that occurrence of E_s is not related to meteoric activity. Experimental evidence is adduced to show that what the previous observers recorded as E_s echoes during meteoric showers, were not really so,, but were merely reflections from meteor trails picked up by the ionospheric recorders. According to these investigations, it is possible to distinguish in ionospheric records the meteor ionization reflections from the E_s ionization reflections. In some of the records true sporadic E's were obtained at the same time as reflections from meteor trails from quite different heights and directions.

(ii) *Thunderstorm and barometric effects.*—It has been suggested by many ionospheric workers that the abnormal increases in Region E critical frequency (which are now identified with E_s) are in some way associated with thunderstorms [61, 207, 208]. Association of E-abnormalities with the isobaric situation near the ground has also been reported by observers in the different parts of the world [203, 204]. According to some other observers, however, evidence of correlation between occurrence of abnormal E and thunderstorm is inconclusive [205, 206]. Notwithstanding these contradictory results, it should be mentioned that increase in the incidence of E_s has been found to occur year after year at Calcutta, India during pre-monsoon (May, June) and monsoon months (July, August) when thunderstorms are very frequent. In what follows we shall discuss two processes by which, according to C. T. R. Wilson, a thundercloud may cause increase of ionization in the region immediately below the E-layer [209].

A thundercloud is charged positively in its upper and negatively in its lower part. Such a cloud has an intense electric field in its interior directed downward. For points situated at a distance, the cloud behaves like an electric doublet. The field in the interior as well as the field outside due to the doublet may cause increase of ionization. First, consider an electron inside the cloud. The electron may be of radioactive origin or may be one produced inside the cloud by local discharges. The intense electric field inside the cloud will drive the electron upwards. Since the gain in energy during its travel in the electric field exceeds the loss of energy due to collision along its path, the energy of the electron will rapidly increase and, as has been shown by Schonland [210], may approach a limiting value of 5×10^9 eV. A stream of such 'run-away' electrons moving upwards past the upper boundary of the cloud will in general be attracted back towards the cloud by the strong upwardly directed field of the doublet. But when a spark discharge occurs the retarding field is destroyed and the spray of 'run-away' electrons moves upward with tremendous velocity. Depending upon their initial velocity these will either be bent down by the earth's magnetic field or reach Region E. In the latter case the run-away electrons would be able to affect the ionization of Region E.

Secondly, the electric field due to the cloud-doublet might be so intense in the neighbourhood of the lower boundary of Region E that an electron accelerated by the field is able to produce electric discharge in this region. We can easily estimate the field of the charged cloud to see if increase of ionization by this process is possible. The electric moments of thunderclouds have been estimated by various workers. In South Africa, Schonland obtains the value 3×10^{16} e.s.u. while in South-East England, Appleton, Watson Watt and Herd [211] obtain the value 2×10^{17} e.s.u. In subtropical regions where the abnormal increase of critical frequency in summer months is very marked the value of the electric moment of thunderclouds may be as great as 5×10^{18} e.s.u. The electric field of such a cloud will be about 6 volts/cm. at a height of 80 km. Free electrons, if present

at such heights, will be acted upon by this field. Now, the pressure at a height of 80 km. is known from various considerations to be of the order 10^{-2} mm. Since the mean free path l of an electron at such pressure is 3·2 cm. approximately, the energy gained by the electron will be about 19 eV. This is just sufficient to ionize by collision the oxygen and nitrogen molecules present in the region concerned. It may also be noted that the free paths of some electrons will be much greater than the mean free path. Such electrons need move in a field of correspondingly lower intensity to acquire the energy necessary for producing ionization.

Besides these two processes a third possible process has been discussed by Bailey and by Healey [212]. The electric discharge of a thundercloud causes radiation of electromagnetic energy (e.g., atmospherics). It is shown that the intensity of such radiation field may be sufficiently strong to produce increase of ionization in the upper atmospheric regions where collisional frequency is less than 10^6 per sec.

(iii) *Solar corpuscles.*—In high latitudes near auroral zones there is another agency, a cosmic one, namely, fast charged particles emitted by the sun, which is known to produce abnormal effects in the ionospheric regions. These will be discussed in sub-section (*f*).

(*c*) Scattering of radio waves by ionospheric irregularities

The irregularities and unevenness in the ionization of the *E*-layer are known to scatter radio waves incident on them. Two types of scattering may be distinguished. First, back-scattering of waves as pierce the ion clouds (on account of the higher wave-frequency) and second, reflective scattering or diffractive reflection of waves (of frequency less than penetration frequency) from the irregularities. The latter type of scattering is known to produce fading of single down-coming signals.

(i) *Back-scattering.*—Careful observations show that the 'skip zone', that is, the region extending from the neighbourhood of a transmitter to the point where the indirect ray first comes down to the earth, is not a zone of complete silence for a powerful transmitter. Due to scattering in the ionosphere steady signal of moderate intensity appears within the zone [213]. The phenomenon was first studied by the Trans-Radio Company of Berlin in collaboration with the British Post Office on wavelengths 14 to 50 m. Eckersley [108, 214] appears to be the first to make a detailed study of the phenomenon and locate the ionospheric region where the scattering takes place. He employed powerful 'pulsed' transmission both omni-directional and beam and observed the scatter echoes with the conventional cathode ray oscilloscope display. The receiver was placed not far from the transmitter. The usual echo pattern obtained with such arrangement is shown in Fig. 76. The ground pulse G is followed by two groups of echoes, one at P and the other at S having a leading edge Q. These groups are usually referred to as *short-scatter* (abbreviation of short-delay

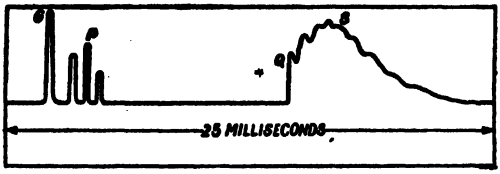


FIG. 76. Echo pattern due to scattering of radiowaves from the ionosphere. G—ground pulse; S—scattered echoes.

scatter) and *long-scatter* (abbreviation of long-delay-scatter). The short-scatter appears intermittently, mostly as discrete peaks for a few seconds at a time at distances of 80–300 km. The long-scatter pattern appears as a persistent, 'boiling' assemblage of peaks at equivalent paths varying from 600 to 2500 km. according to the ionospheric condition. The origin of the short-scatter is now universally recognized as back-scatter from irregularities near E-layer (see Fig. 77). Regarding the long-scatter, the

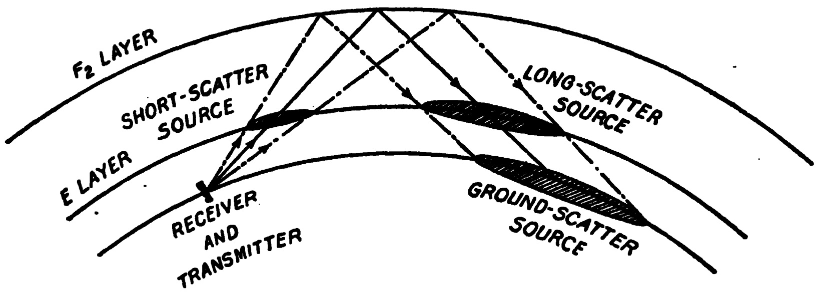


FIG. 77.—Illustrating sources of back-scattered echoes. Ionospheric scattering occurs at irregularities at the E-region. Scattered echoes of long-delay are generally ascribed to back-scattering, at ground, of waves reflected from the F-layer, though, there may also be back-scattering by the E-layer irregularities.

consensus of opinion is that it is due to back-scatter from ground returning to the receiver *via* F-layer (Fig. 77). However, observations by Silberstein (operating on 13·66 Mc./s. with peak-pulse input 770 kw.) show that though most of the long-scatter is due to scattering at the ground, there are occasions when back-scattering from the more distant point of the E-layer is also observed [215]. According to Eckersley and Farmer [108] each echo is due to superposition of scatterings from a number of ion clouds and not due to scattering from a single cloud. This conclusion is in accordance with the hypothesis that the ionization close to the E-layer is 'patchy'.

Observations at Lindau, Germany, by Dieminger during 1947-48 using high power transmitters (10–20 kw. and 100 kw.) also showed that the long-scatter echoes are due to back ground-scattering of down-coming waves *via* F-layer and not to scattering in the F-layer [216]. Scatterings from the E and F_1 layers were also observed. The reason why no scattering is obtained from the F-layer is that on account of large mean path and high temperature the rate of diffusion is much larger than in the E-layer. Any

local irregularity or 'patch' formed has little chance of persisting as such for any length of time, being diffused away soon after they are formed.

The characteristic grouping of the long-scatter echoes is ascribed to two focussing effects [217]. One, 'angle-focussing' effect caused by the earth-curvature of the F-layer, the other, 'time-delay focussing' caused by the ground back-scatter signals from a wide range of ground distance tending to arrive (due to the refraction characteristic of a parabolic gradient of ionization) within a narrow interval, producing a 'bunching' in time. Peterson [217] has studied theoretically the two focussing effects and has shown that combination of the two yields an intensity distribution which agrees fairly well with the observed intensity pattern of long-scatter echoes, namely, a maximum in the distribution and a leading edge.

Another type of echo, called G-scatter is also reported on frequencies much higher than the F-layer critical frequency [216, 218]. They disappear gradually on increasing the frequency without much alteration in height and appear to come from virtual heights 800 to 1200 km. According to Rivault [218] this scattering is associated with magnetically disturbed and meteoric shower periods. Dieminger [216] is of opinion that this type of scattering is similar to scattering as observed regularly at Huancayo [219] and may be attributed to 'simultaneous reflections on normal ionospheric layers and a disturbed belt to the north of the observing station.'

(ii) *Reflective scattering.*—We now consider the other type of scattering, which may be called reflective scattering or diffractive reflection. Such scattering manifests itself by the fading of single down-coming waves.

It is common experience that fading of indirectly received signals occurs even when there is only one down-coming wave (i.e. there are no multiple waves due to multiple hops or to magnetic splitting) and there is no ground wave of appreciable strength. Fading of such single down-coming wave is ascribed to irregular reflections from a wavy uneven 'bottom' of an ionized layer. Due to irregular reflections the single wave received at a point cannot be considered as a beam reflected from a smooth mirror-like surface, but as a cone of rays (divergent or convergent) returned by diffuse scattering. The intensity of the wave is different at different neighbouring points and there is a sort of diffraction pattern of the received signal on the ground. If the ionospheric irregularities are fixed in space and time, then the resulting diffraction pattern on the ground is also fixed in space and time and there is no fading in a stationary receiver. Two extreme cases may now be imagined. The irregularities—the scattering centres—instead of being fixed, may be imagined to be moving as a whole in an orderly fashion in a particular direction, say horizontally. The diffraction pattern on the ground will then be moving past the fixed receiver and there will be fluctuations of the signal strength or fading. We may also imagine that the scattering centres are moving in a completely random fashion. In this case the diffraction pattern will be continuously changing and there will again be fading of the received signal. In this case, however, the fading is complicated by the fact that there will be Doppler shifts in the frequency

of the down-coming waves due to the motions of the scattering centres having components in the line of sight [220].

It is possible by statistical analysis of fading records in spaced receivers to distinguish between the two types of fading [221]. There is also the possibility that the observed signal intensity at any instant is not wholly due to irregular scattering, but partly to specular reflection. The percentage of such contribution may also be determined by statistical analysis [222]. Significant departures from the observed probability distribution of amplitude from the Rayleigh law (which applies to random distribution of the scattering centres) have been observed by workers in India [223, 224, 224*a*]. Such departures have been ascribed to the existence of (either or both) steady drift and specular reflection.

Statistical analysis of the fading records made in three receivers placed at the corners of a right-angled triangle also enables one to determine, (*a*) the steady velocity of the 'diffraction' pattern over the ground, and, hence, the velocity of the scattering centres; (*b*) the rate at which the diffraction pattern is altering as it moves; (*c*) the size of the irregularities in the pattern [225]. In Sec. 14 we shall describe in some detail the experimental arrangement as adopted by S. N. Mitra to determine the horizontal velocity of the scattering centres. It is interesting to note in this connection that sporadic *E*, as recorded in collaboration with amateurs all over the U.S.A., showed that the reflections came from the localized patches which moved in definite directions [219].

(iii) *Theoretical study.*—The problem of scattering by ionospheric irregularity has also been studied theoretically. Scattering by a single region of inhomogeneity—a long thin cylinder of ionization as produced in a meteor trail—has been studied by Blackett and Lovell [227] and has been referred to in Chap. IIIB, Sec. 4.

Booker has examined closely the problem of scattering by the sporadic *E*-region and has developed a theory based on some earlier work of Booker and Gordon [228, 229]. The theory can also be applied to the irregularities in ionosphere produced during auroral activity and during magnetic storm. The most important deduction of the theory is that the wavelength at which E_s echoes disappear is determined by the scale l of the fine structure as used in the theory of turbulence (critical wavelength $\lambda = 4\pi l$) rather than by the electron concentration. Starting with longer waves it is found that as the wavelength is decreased, the scattering is omni-directional and the echo-strength independent of wavelength down to about wavelength $4\pi l$. For shorter wavelengths the scattering is predominantly in the forward direction and the electric field strength of the backward scatter echo decreases inversely as the square of the wavelength. There is thus a sudden decrease in the strength of the scatter echo (backward) round the wavelength $4\pi l$. It may be mentioned in this connection that according to Rawer's observations E_s shows a marked change of reflection coefficient with frequency, which is not expected from regular reflection from a continuous layer [230].

In order that the deductions from the theory may fit the observations it is necessary that the statistical departures of the electron number density should vary from a few per cent to about 30 per cent from the mean. However, it is found that, fortuitously, the MUF for communication to a distance *via* E_s is very similar to what is expected from the usual calculations from the maximum electron number density instead of the fine structure scale l.

According to Booker scattering in the E-region may produce ionospheric twinkling in the intensities of radio signals as received at the ground from gallactic sources.

(*d*) Sudden Ionospheric Disturbance (S.I.D.)—Radio Fade-out

Sudden and intense increase of ionization in the ionosphere is found to occur simultaneously with solar flares—appearance of bright spots on the solar disc. The increase causes weakening and, in the extreme cases, complete 'fade-out' of radio signals over the sun-lit portion of the hemisphere on short and medium wavelength. The fade-out may last from a few minutes to an hour or so. The main features of the fade-out phenomenon were first ascertained by Mögel between 1927–30 and attention to it was redirected by Dellinger in 1935 [213, 231]. The phenomenon is thus also known as the Mögel-Dellinger effect.

During the fade-out period the values of terrestrial magnetic elements as well as those of the earth-currents are found to undergo sudden changes. The word 'fade-out' refers only to short and medium waves; for the long waves the enhanced ionization causes an increase, instead of a decrease, in the signal strength. It may be recalled here that a possible relation between bright solar eruption and sudden variation of terrestrial magnetism was hinted long ago by Carrington when on September 1, 1859, at 11 a.m., he observed visually patches of intensely bright and white light on the solar disc while watching sunspots. Magnetogram records at the Kew Observatory showed simultaneously sudden jumps in all the three magnetic elements. The observation excited considerable interest at that time but was later neglected as it was thought to be a spurious coincidence of solar eruption and magnetic disturbance.

The most natural explanation of the radio fade-out is that the bright chromospheric eruptions of the sun send out extreme ultraviolet radiation which penetrates into the lowermost levels of the ionosphere and causes increase of ionization in the region where collisional frequency is large. This region thus becomes a strong absorber of short and medium radio waves. The increased conductivity following the increased ionization enhances the dynamo effect [Chapter VII, Sec. 5] and thus intensifies the diurnal variation of the magnetic elements.

The simultaneity of solar, magnetic, earth-current and radio effects is pictured in Fig. 78 (Plate III) which depicts records of ionospheric, magnetic and earth-current measurements at Huancayo and spectroheliograms taken at Mount Wilson Observatory on April 8, 1936. The

phenomena depicted in the figure may perhaps be best described in the words of McNish [232]:

'Routine observations with the spectroheliosocope were in progress at the Huancayo Observatory at the Carnegie Institution's Department of Terrestrial Magnetism at this time. The observer on duty noticed a remarkable brightening of the *H*-alpha light in the region of a large sunspot, beginning at 16 h. 45 m. G.M.T. This eruption was accompanied by certain terrestrial phenomena as shown in Fig. 78 (Plate III). The spectro-heliograms shown in the figure were made at Mount Wilson before the eruption, while the eruption was at its maximum, and after the eruption had subsided. They show the brightening of the sunspot, although high contrast of the negatives used permits little discrimination regarding intensity. The observers at Huancayo reported that from the viewpoint of extent and intensity this was the most spectacular eruption which had been witnessed at the observatory since beginning routine observations with the spectro-helioscope a year before.

'Ignorant of the solar fireworks in progress, the radio operator at the station, who was engaged at the time in making measurements on the ionosphere, found it impossible to receive reflections on any frequency and began to search for defects in his equipment. Finding none, observations were resumed. The disappearance of these echoes and their subsequent return 55 minutes later are illustrated at the top of Fig. 78 (Plate III). The sudden changes in the terrestrial magnetic variations and in the earth-currents which are continuously registered at the Observatory are conspicuous in the records at the bottom of the figure. It may be noted that all the terrestrial effects began at 16 h. 46 m., one minute after brightening became perceptible in the sunspot-region.'

We shall now discuss the various aspects of the phenomena in some detail.

(i) *Radio effects.*—The most spectacular aspect of the radio effects is, as already mentioned, sudden cessation of wireless traffic on the sun-lit portion of the earth on short and medium waves. We give below results of other observations on radio wave propagation during solar flares.

It has been found that during a solar flare there is, on all observed wavelengths, a lowering of the virtual height of reflection. This is revealed by a sudden change of phase of the sky wave. This phase change (usually called sudden-phase-anomaly—S.P.A.), has been studied in great detail for very-long-waves (16 kc./s.) in England [232*a*, 233]. The average lowering of the virtual height of reflection is 4 km., but may even be 15 km. It has been found that the S.P.A. is very sensitive to the occurrence of solar flares and, in fact, provides a much better index of such occurrences than the 'fade-out' which is observed only when the flare is intense. The anomaly lasts over a time of several minutes and is followed by a recovery which takes about three times as long.

Unlike the phase anomaly effect, the effect on the intensity of the reflected wave during a solar flare is not the same for all wavelengths. It

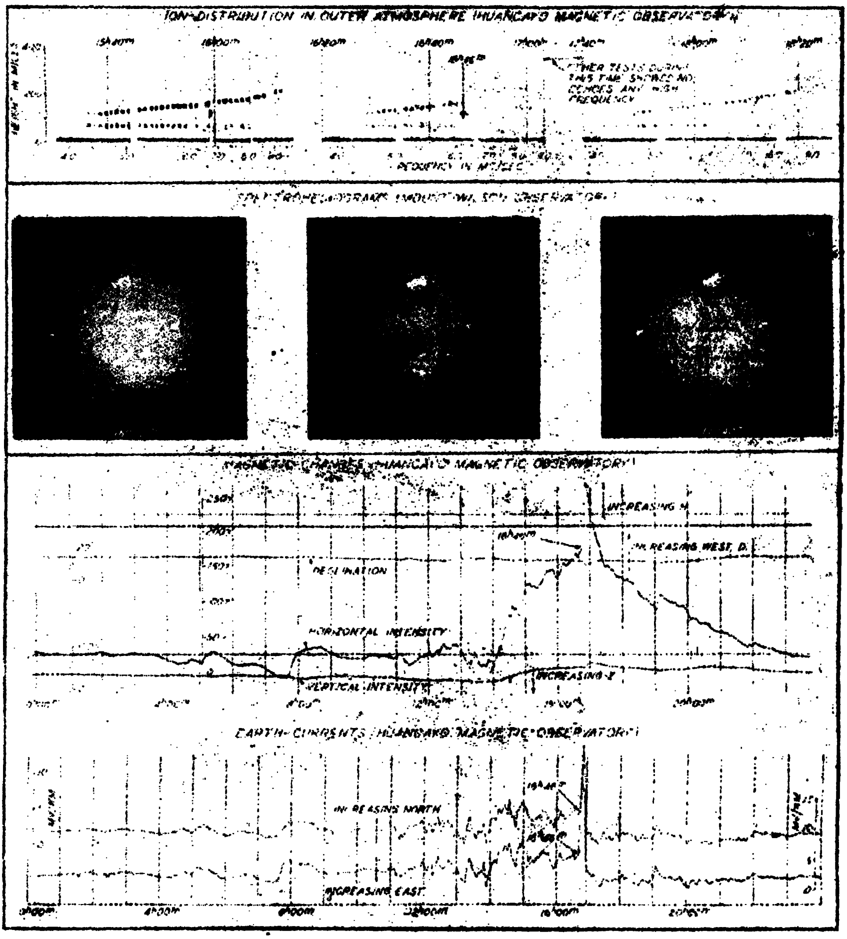


FIG. 78. Radio fade-out and associated phenomena. Note the simultaneity of break in ionospheric record and onset of disturbances in terrestrial magnetic and earth-current records associated with bright solar eruption. Records made on April 8, 1936, 16 h. 46 m. G.M.T.

The increase in the value of the terrestrial magnetic field is ascribed to enhanced intensity of the *Sq* type of upper atmospheric electric current system. [See Chapter VII, Sec. 3(c)]. (After McNish.)

FIG. 83.—Noctilucent clouds as observed from Drobak, Norway on the night of 27-28 July, 1909 at 23 h. 1 m. The height of these particular clouds were not measured. But measurements carried out elsewhere in Norway show that the average height is 82 km. lying between 74 and 92 km. (After Störmer.)

depends both upon the wavelength and on the angle of incidence on the ionosphere. For very-long-waves (16 kc./s.) at moderately steep incidence the intensity of the sky-wave is hardly, if at all, affected [233, 146]. If, however, the incidence be very oblique then according to Bureau's observation there is an increase in the field strength [234]. It has also been found that atmospherics (which are known to possess a quasi-periodicity of the order of a thousand cycle per second) of distant origin increase greatly in number during a fade-out [234].

For long waves (frequency of the order 100 kc./s.) it is found that for steep angle of incidence there is a decrease in intensity [235]. But, for oblique incidence (transmission over 1000 km.) there is considerable increase in the reflection coefficient [235]. Measurements show that in such cases the intensity of the sky wave may be as much as one-half of that of the ground wave, though ordinarily its value is only about one-fifth [236, 237]. The interference of the ground wave with the sky wave of enhanced intensity produces fluctuations in signal strength during a solar flare, as the effective height of reflection falls and rises decreasing and increasing the path difference [238].

On higher frequency (2 Mc./s.) for both vertical and oblique incidence there is a decrease in signal intensity, i.e., the fade-out effect [239].

Fig. 79 shows the simultaneity of sudden decrease in short wave signal

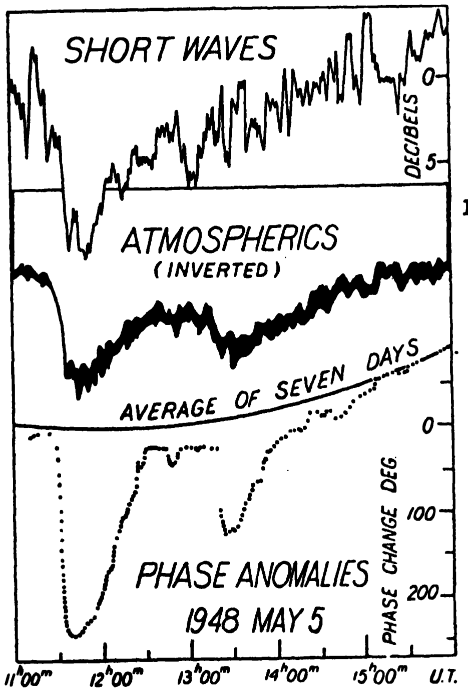


FIG. 79.—Illustrating the simultaneity of sudden decrease in the signal strength of short wave (Leipzig, 9·732 Mc./s.) observed at Bagneux, of increase in the number of atmospherics (recorded on 27 kc./s.) at Poitiers and of sudden phase anomaly of down-coming very-long-waves (16 kc./s. from G.B.R. Rugby) observed at Cambridge, England. (After Bracewell and Straker.)

strength (Leipzig 9·732 Mc./s.) observed at Bagneux, of increase in numbers of atmospherics (recorded on 27 kc./s. at Poitiers) and sudden phase anomaly

of down-coming very-long-wave (16 kc./s. from G.B.R. Rugby) observed at Cambridge, England [233].

Radio effects in the F_2-layer during a fade-out are not so marked as in the lower regions. A slight diminution in electron density accompanied by an increase in the virtual height has been reported [240]. Further, the change in the F_2-region begins to occur almost simultaneously with the chromospheric eruption, though, the fade-out effect usually occurs after the commencement of the same. According to Becker and Dieminger [241] the effective recombination coefficient in the F_2-region during fade-out may be as large as $1{\cdot}9\times10^{-10}$cm.^{3}/s. The ion-production in the beginning of the effect may be double that of the normal ionizing rate.

Ionospheric effects of solar flares are regularly reported by the Research Laboratory of Electronics, Chalmers University of Technology, Gothenberg, Sweden, National Research Council, Ottawa, Canada, and the Central Radio Propagation Laboratories, Washington.

(ii) *Origin of the sudden ionospheric disturbances.*—There is little doubt that the sudden ionospheric disturbances are caused by intense ultraviolet radiation (emitted during solar flares) penetrating deep into the atmosphere and increasing the ionization in the lower absorbing regions of the ionosphere. It is clear that the radiation must be such as to be able, on the one hand, to ionize one or other of the atmospheric constituents in this region and, on the other, must have sufficiently small absorption coefficient to reach such low regions as 60–70 km.

Regarding the probable wavelength of the active radiation we first note the following. During a solar flare the visible Balmer line of hydrogen $H\alpha$ is found to be emitted with great intensity. Measurements also show that the width of the $H\alpha$ line during a flare varies in a characteristic manner which bears a close relation to the intensity of the flare [242, 243]. Also, it has been found that the behaviour of the $H\alpha$ line width during a flare shows great similarity to the time-variation of the phase changes of the down-coming very-long-wave, i.e., of the virtual height of reflection, which, in a sense, may be taken as a measure of the variation of the ionization density during a flare [233].

Now, it is reasonable to suppose that during a flare, along with $H\alpha$ line, the Lyman lines of hydrogen in the extreme ultraviolet, $L\alpha$—1215 A., $L\beta$—1026 A., $L\gamma$—973 A., etc., are also emitted with great intensity and that their intensities vary in a manner similar to that of $H\alpha$. Hence, in view of the close correspondence between the intensity variation of the $H\alpha$ line and of ionization density as noted above, one may reasonably suppose that one or other of the Lyman lines is the active ionizing agent during a solar flare. The wavelength of the actual line must, of course, be below the critical ionizing wavelength of the atmospheric gases and must also, at the same time, have sufficiently low absorption coefficient to enable it to penetrate into the atmosphere down to 60–80 km.

Regarding the atmospheric constituent affected, it has been suggested that O_2 may be the active gas [129*a*]. Since the ionization potential of O_2

is 12·2 eV (the corresponding wavelength being 1019 A.) it is clear that neither L_α nor L_β can be effective. L_γ (973 A.) can therefore be the active radiation. And, as its absorption coefficient is relatively feeble it will be completely absorbed only after it has penetrated down to a considerable depth below region E. Calculation by Mitra, Bhar and Ghosh [129a] on the assumption that the absorption coefficient of O_2 for this radiation is only one-thousandth of that of λ 770 (corresponding to the second ionization potential) shows that the maximum ionization due to this radiation occurs at a height of about 55 km. With the assumption of a higher absorption coefficient the maximum would necessarily be higher.

It may be noted that records of sudden ionospheric disturbances (as revealed by increase in the number of atmospherics) show that their frequency increases and decreases with the 11-year solar cycle [234].

(e) Ionospheric storms

Sudden changes in the F_2-region of the ionosphere (consisting generally of a depression of the ionization density and increase of the virtual height) associated with terrestrial magnetic disturbances are known as ionospheric storms. These storm-time changes may be contrasted with the S.I.D. discussed above in which the lowermost regions of the ionosphere are mainly affected and which are associated with solar flares. The ionospheric storm effects are dependent on the latitude—magnetic rather than geographic. The following remarks regarding the latitude distribution of the ionospheric storm characteristics are from analysis of world-data made by Appleton and Piggott [149].

In high latitude the F_2-layer critical frequency is depressed with the onset of geomagnetic disturbance. As soon as the violent perturbations of the geomagnetic field subside the critical frequency recovers its normal value. These effects are characteristic of the auroral-type conditions and may be designated as *Q.R. negative phase*—Q.R. standing for 'quick recovery'. (Other effects on ionospheric conditions at auroral latitudes, associated with geomagnetic disturbances will be discussed in the next sub-section.)

In temperate latitudes there is a different type of ionospheric storm manifestation. This is depicted in Fig. 80. In drawing the curve, days of major depression of the ordinary ray critical frequency ($f^\circ F_2$) associated with geomagnetic disturbance are taken as the reference (zero) days. For a number of days before and after zero days, as also for the zero days, the average departures of $f^\circ F_2$ from the months' mean noon values (denoted by $\Delta f^\circ F_2$) are plotted. It will be noticed that $f^\circ F_2$ has been significantly above the average for two days. This positive-negative sequence may be called a 'bi-phase' disturbance. The slow process of recovery, in contra-distinction to the quick one mentioned above, may be called *S.R. negative phase*—S.R. standing for slow recovery.

In an equatorial station there are comparatively few days in which there is a sudden drop in $f^\circ F_2$ value. If, however, storm days of the

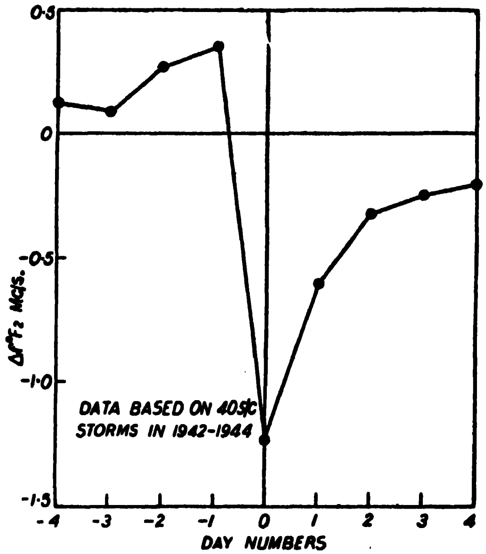


FIG. 80.—Illustrating variation of $\Delta f^{\circ}F_2$ (average departures of $f^{\circ}F_2$ from the month's mean noon value) with the day-numbers of magnetic disturbances. (The data are for temperate latitudes.) The days of major depression of the ordinary-ray critical frequency ($f^{\circ}F_2$) associated with geomagnetic disturbances are taken as the reference zero days. (After Appleton and Piggott.)

temperate latitudes are selected, and, if for these days superposed-epoch curve (with the equatorial station data) is drawn, then it is found that there is a well-marked positive phase, coinciding roughly in time with the negative phase in temperate zone.

As already mentioned, the type of storm which a given station experiences is correlated more to the magnetic than to the geographic latitude. But, even for the former, the connection is rough. For example, Washington (magnetic latitude 50° N.) behaves mostly like an auroral zone type station, but Leningrad (magnetic latitude 49° N.) is predominantly non-auroral.

In Chap. VII we shall refer to magnetic storms of a world-wide character which commence suddenly. It has been found that on many such occasions a rapid depression of $f^{\circ}F_2$ occurs followed by a short positive phase. This is a characteristic of only temperate and higher latitude stations.

According to Appleton and Piggott ionospheric storm phenomena contribute substantially to the general variability of the F_2-layer.

An interesting effect on the E-region during the severe magnetic storm of May 11–13, 1949, which has been reported from Stanford, California, U.S.A. [192] may be mentioned here. It was observed that reflections from the E-region (on 850 kc./s.) suddenly disappeared for periods lasting 10–15 minutes. Two such 'voids' in the h'–t trace occurred in course of the storm. The first, during the 'mild phase' about 21 hours before the 'sudden commencement' and the second during the 'turbulent phase'. On both occasions, following the 'voids', the reflecting E-layer, reappeared on the h'–t trace in their original form. It appears, as if, there are present during ionospheric storms some agency which can affect E-region ionization without disturbing the F-region and, also, some other agency which reforms the E-region in course of a few minutes.

(*f*) Abnormalities near auroral zones

Observations show that ionospheric conditions near auroral zones (round magnetic poles) differ markedly from those at low and temperate latitudes. This is ascribed to two causes: First, the diurnal and seasonal variation of the illumination of the high atmosphere by solar rays is quite different in such regions from that at lower latitudes; for periods extending over months the overhead atmosphere may be totally devoid of solar rays or illuminated only by oblique rays from below the horizon (see Appendix, Sec. 1). Secondly, the incidence of charged solar corpuscles, as cause aurorae and magnetic storms, affect the ionization of polar regions.

Some general features of the ionosphere near auroral latitudes may first be mentioned. In such latitudes, even the calmest days appear as disturbed in comparison to magnetically quiet days at lower latitudes. Abnormal condition of the ionosphere may thus be regarded as almost the normal feature of the ionosphere. Complete cessation or 'black out' of echoes on all wave frequencies is of common occurrence.

The remarks on ionospheric conditions in auroral latitudes that are to follow are taken mostly from observations made at three high latitude stations: Tromsö (69° 39′ N., 18° 56′ E.), East Land (80° 23′ N., 19° 31′ E.) and College, Alaska (64° 41′ N., 148° 25′ W.) [120, 245, 246].

(i) *Diurnal and seasonal variation of ionization on magnetically quiet days.*—During the period when the upper atmosphere is illuminated (e.g. during summer solstice), the average diurnal variations, for magnetically quiet days, of E, F_1 and F_2-regions are, in general, similar, with certain minor differences, to those observed at lower latitudes. During the dark period the diurnal variation of the F-region can only be observed. It has a minimum soon after midnight and maximum just before local noon. F-echoes during winter nights are attributed to formation of sporadic F-layers, it being suggested that the agency which produces F-echoes in the polar nights also cause increase of F-critical frequency in winter nights in middle latitudes [245].

The average seasonal variations of the E and F_1 layers for the magnetically quiet days appear to follow approximately Chapman's relation (Eq. 129). In Fig. 81 the continuous line curves depict the seasonal variations of ionization of the E and F_1 regions at Tromsö as expected according to Eq. (129) [120]. It will be noticed that the ionization value drops and rises abruptly in the months of November and January respectively marking the beginning and the end of the period of darkness. The dots and circles, representing the observed values of the critical frequency for the E and F_1 regions respectively are seen to group round the theoretical curves showing that the average seasonal variation of ionization on magnetically quiet days follows the Chapman relation. It will be noticed, however, that even during the period of total darkness, when, according to theory, the ionization should disappear completely, sporadic E-echoes have been obtained and that the critical frequency has sometimes been as high as 1·6 Mc./s.

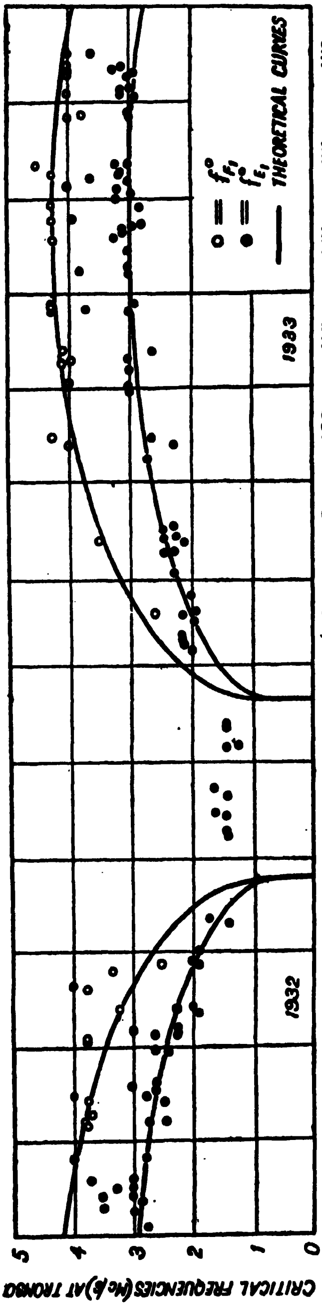


Fig. 81. Observed and calculated values of ordinary ray critical frequencies at noon at Tromsö, Norway. (After Appleton, Naismith and Ingram.)

No pronounced seasonal variation the F_2 (noon)-ionization is observed owing to the extreme variability of the ionization. It should be noted, however, that in common with the seasonal variation of F_2 at lower latitudes, two maxima—one at the vernal and the other at the autumnal equinox—are observed. The Tromsö observations during the polar year 1932-33 show that the sporadic E might be so persistent during midsummer that the diurnal variation of F-ionization cannot be observed [120, 247].

(ii) ***Magnetic activity and ionospheric reflection conditions.***—Ionospheric reflection conditions in auroral latitudes, associated with magnetic activity, may be summarized thus [120]:

The echoes are complex in nature and are variable in intensity. Just before a strong magnetic disturbance the complexity increases to such an extent that it becomes practically impossible to distinguish between E- and F-echoes, there being only a blurred pattern indicating reflections from heights of 100 km. and upwards.

At every magnetic disturbance the reflection coefficient becomes abnormally low. If the storm is intense, the reflection coefficient falls to such a small value that there is 'black-out' of the echoes on the usual range of wave frequencies.

Critical frequencies of the E, F_1 and F_2 regions show, in general, an inverse relation with the magnetic character figure, i.e. the higher the character figure the lower the critical frequency. Such inverse correlation has also been observed in regard to the winter thickness of the F-layer in high latitude stations [244]. Weak magnetic disturbances are always found to be associated with sporadic E-reflections.

The sequence of the effects mentioned above may be thus described

when a storm is in progress: Just before the beginning of a storm, the echoes become diffuse and scattered. *F*-critical frequency begins to decrease while sporadic *E*-echoes appear. The intensity of the echoes, however, decreases rapidly. With the progress of the storm E_s persists on higher and higher frequencies and *F*-reflection ceases. When the perturbing magnetic vector increases to its maximum value the black-out' condition sets in and usually lasts for nearly an hour afterwards. There is almost invariably a black-out condition on the day following a night storm even though the disturbing magnetic element has resumed its normal value.

According to observations by Wells [245] there is a direct relationship between occurrence of the 'bay' type of magnetic disturbance (see Chap. VII, Sec. 8(*a*)) and increase of ionospheric absorption in auroral regions leading to radio 'black-out'. Magnetic bays frequently run in series of two to five days, the time interval being close to 24 hours. These recurrences of magnetic bays are always accompanied by the radio black-out phenomena (see Fig. 82). There is also a close correspondence between the occurrence of polar black-outs and polar sporadic *E*-ionization.

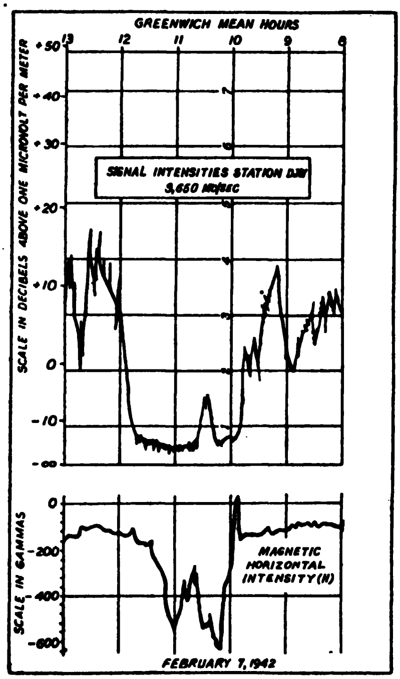


FIG. 82.—Illustrating how, at auroral latitudes, occurrence of magnetic bays is accompanied by the decrease of signal intensity (radio black-out). (After Wells.)

According to analysis carried out by Nagata the electron concentrations in the *E* and *D* layers in the auroral zone are increased during periods of magnetic disturbance by a few ten-times of those during calm period [141].

The magnetically disturbed condition of the ionosphere in auroral latitudes has its effect on long distance radio transmission. If the transmission path (the great circle joining the transmitting and the receiving stations) pass through the magnetically disturbed zone—a circle of radius about 30° round the magnetic poles—the service is found to be very poor. In fact, it has been observed that the further the transmission path lies from the magnetically disturbed zone the better is the reception.

14. WINDS, TIDES AND TRAVELLING DISTURBANCES IN THE IONOSPHERE

(*a*) Introduction

In Chapter II we have shown how motions due to tidal forces extend into the high regions of the atmosphere. As such, the ionospheric regions

participate in these motions [248]. It is, however, now generally recognized that in the high atmospheric regions, there exist, besides motions due to tides, motions due to pressure gradients caused by temperature inequalities. It has, in fact, been suggested that there are, resembling the general circulation in the lower atmosphere, circulatory motions between the E and F layers [249]. However, such thermal circulatory motions in the high atmospheric regions still remain to be explored with the same amount of detail as in the tropospheric regions.

In the present section we shall first discuss the evidence and the measurement of winds near the E-region. (It is not yet established if these winds are caused by tidal forces or by thermal inequalities.) Effects of tidal motions on the variations of the ionospheric parameters h' (height of the bottom of the layer), h_{max} (height of maximum concentration of the layer) and f_0 (vertical incidence critical frequency of the layer), will then be discussed. Such effects are particularly noticeable in the F_2-region where, on account of the rarity of collisions, the motions of the ions and electrons, as would otherwise be produced by the tidal forces alone, are profoundly modified by the terrestrial magnetic field. Finally, we shall discuss the case of travelling ionospheric disturbances in which some characteristic feature of ionization (e.g. a characteristic hump in the h'-t curve) appears to travel as a disturbance wave from over one observing station to another.

(*b*) Winds and wind velocity measurement

The most direct evidence of winds in the ionospheric regions are those furnished by the motions of the so-called noctilucent clouds and the drift of long enduring meteor trails.

The noctilucent clouds are observed in high latitudes some hours after sunset or before sunrise being illuminated by the rays of the sun from below the horizon [250] (see Fig. 83, Plate IV, facing p. 327). They were first recorded by Jesse in 1890 [251] and have since been observed on many occasions in northern and southern latitudes, always in the summer months in the place of observation [251*a*]. The origin of these clouds is still uncertain. According to one hypothesis they are composed of ice crystals (and, hence, a low temperature is postulated in the region of occurrence); according to another, they are due to the cosmic dust produced by violent volcanic eruptions like the Krakatoa or, by fall of large meteorites, e.g. the Great Siberian Meteorite (Chap. II, Sec. 4). The heights of the clouds have been measured by observing them from two stations by the method similar to that of the auroral height measurement (see Chap. VIII, Sec. 2(*b*)). The heights are found to lie within a rather narrow range between 74 km. and 92 km., the average being 82 km., i.e. somewhat below the E-layer. The velocities vary within wide ranges. According to the earlier measurements of Jesse the velocities varied between 100 and 300 m./s. The directions were mostly from ENE. According to measurements of Störmer made in 1932 in

Southern Norway, the clouds were found to move from NNE with velocities between 44 and 55 m /s. [250].

As already mentioned visual observations on meteor trails give evidence of high wind velocity in the region 80–120 km. Method of measuring the wind velocity by observations on the meteor trails by the radar technique (see Chap. IIIB, Sec. 5) has been developed [251*b*].

A simple ionospheric evidence of wind in the E-region is the similarity, with a time lag, in the echo-patterns recorded at two receivers situated some distance apart [252, 252*a*]. It is natural to suppose that the similarity is due to some characteristic ionospheric irregularity ('patches' or 'clouds' of intense ionization as produce E_s) travelling within the time lag interval from over one station to the other. In what follows we shall describe briefly an experimental method adopted by S. N. Mitra for measuring wind velocity in the E-region from observations on similarity of fading of a single down-coming wave at spaced receivers [252*b*].

Let a radio wave be incident on the region of ionospheric clouds which are assumed to lie in a horizontal plane. The intensity of the reflected radio wave as received at a ground station is determined by the effect of the superposition of the wavelets scattered from the irregularly distributed clouds. If the clouds are at rest, then the received signal intensity, though varying from point to point, will be constant in time at a given point. The clouds may, however, be in motion, and, we may imagine two extreme cases : Firstly, a steady drift of the clouds as a whole in some direction, and, secondly, a wholly irregular random motion amongst themselves. (There may, of course, be combinations of the two motions.) In both the cases there will be fluctuation of signal intensity or fading as recorded in a receiver.

For the first case, the fading will be due to the diffraction pattern sweeping past the receiver. Simple consideration shows that the velocity of the sweep will be twice the velocity (v meters/sec.) of the scattering clouds, so that, if there are two receivers, distance d meters apart in the direction of motion, the fading pattern as observed with the first receiver will be repeated in the second receiver after a time interval $d/2v$ seconds.

For the second case, the fading will be due to random scattering by the clouds in irregular motion. The fading patterns in the two receivers, in this case (provided the two receivers are sufficiently separated) will be wholly unrelated to one another.

S. N. Mitra has studied both types of fading with spaced receivers and from the first deduced wind velocity in the scattering E-region of the ionosphere. He used a pulsed transmitter working on 4 Mc./s. in conjunction with three receivers R_1, R_2 and R_3 placed at the corners of a right angled triangle with the arm R_1 R_2 pointing north and the arm R_1 R_3 pointing west. The outputs of the three receivers are brought to a central recording point so that simultaneous records of the fading of a single pulse reflected from the ionosphere are made. If the ion clouds be moving horizontally at a steady rate and if there is no turbulent motion, then the three receivers

will record similar but displaced fading curves. From the time-shift between the fading curves of the receivers R_1 and R_2 and from that of the receivers R_1 and R_3 the north and the west components of the motion are obtained. Hence, the direction and magnitude of the velocity of the scattering clouds are determined.

In the experiments which were carried out at Cambridge, England, the velocities were found to vary between 20 and 110 m./s., the most frequent value being 50 m./sec. The most frequent direction of the wind was found to be towards north-west. From a few 24-hour observations, a semi-diurnal variation of the direction was obtained. But from the meagre experimental results, close agreement between the observed variation and the theoretical variation (direction varying nearly uniformly) was not obtained.

These results may be compared with those obtained by other observers. Meek in Canada obtained the value of about 100 m./s. [252*a*]. Pawsey in England recorded value of 170 m./s. [253]. The value obtained by the Japanese workers is of the order 70 m./s. in night-time, southwards and also probably in daytime, northward [252*c*]. This order of velocity is also suggested from consideration of circulatory motion in the ionosphere [249]. The velocity as deduced from the tidal pressure oscillation theory is of the order 65 m./s. (see Chap. II, Sec. 4).

It may be noted in this connection that some workers (e.g. Munro [253*a*], Beynon [254], Meek [252*a*]) have reported motion in the F-region from their observations. The possible nature and origin of these motions will be discussed in sub-section (*c*) to follow.

Measurements on the wholly random motion of the irregularities due to turbulence have also been made. The fading curves for such cases, as recorded at the spaced receivers are necessarily uncorrelated. Such fading curves have been analyzed statistically [222] after the method of analyzing random noise when passed through a Gaussian band-pass filter. It is assumed that the power spectrum density of the down coming wave is of Gaussian shape, because the velocity distribution of the irregularities is also assumed to be Gaussian in the line of sight. The analysis leads to the evaluation of the r.m.s. velocity of the irregularities in the line of sight. This is found to be 2 to 3 meters per second. No explanation has been offered for the existence of such low value of random velocities of the irregularities in the ionosphere. McNicol [255] has also reported similar observations on the random velocity taking oblique incidence transmission into account.

(*c*) Tidal Effects in the Ionosphere

As the ionospheric layers participate in the tidal motions of the high atmosphere, it is natural to expect that ionospheric records, if analyzed, would reveal the effects of such motions. This, in fact, is what has been found. Analysis of the daily records of h', h_{max} and f° has shown that their values oscillate with the solar and lunar tides. The motions of the

ions and electrons as produced by the tidal motions of the air are, as already indicated, not simple on account of the influence of the terrestrial magnetic field. The observed variations of ionospheric parameters are, therefore, not always easy to interpret in terms of tidal motions.

A theory of tidal oscillations in the ionosphere, the so-called *electrodynamical theory* in which the effect of the terrestrial magnetic field has been taken into account, has been developed by Martyn [248, 257, 258, 259, 260]. According to the theory, many of the observed anomalies in the F_2-region are traceable to the effects of the electrodynamical action.

In what follows we shall first describe briefly the evidences of tidal effects that have been deduced from analysis of ionospheric data. A brief account of the electrodynamical theory and its consequences will then follow.

The principal ionospheric parameters recorded are h', h_{max} and f°, for regions E, F_1 and F_2. Hourly or bi-hourly measurements of each of these data for long periods are necessary for deducing the tidal variations. The solar tidal variations are usually much larger than the lunar ones but are more difficult to isolate. This is because the solar diurnal ionospheric variation has a strong non-tidal 12-hourly harmonic due to the ionizing effect of solar rays of 24-hour fundamental periodicity. Lunar ionospheric tidal effects, though of much smaller amplitude, are free from these complications and are easier to deduce.

(i) *Lunar tidal effects.*—The methods used for the determination of the lunar barometric tide and the lunar magnetic variations (Chap. VII, Sec. 4) may also be employed for deriving the lunar tidal variations of ionospheric characteristics. The solar diurnal variation is first removed from the data. They are then rearranged in lunar time reckoned from the lower transit. The data for each calendar month are arranged in lunar time on one sheet and added up. Selected groups of months are then subjected to harmonic analysis.

As an instance of the analysis of ionospheric data for deducing tidal variations, we may refer here to the work of Appleton and Weekes (261)—the earliest work of its kind—on the lunar semi-diurnal variation of $h'E$ at Slough, England. These authors worked with the data of eleven different periods (in the years 1937-38) each period covering 12–14 days. The height measurements were carried out on 1·8 Mc./s. As this frequency was below three-quarters of the critical frequency, the true height was practically the same as the equivalent height. The hourly means for the whole period of observations were plotted and a smooth curve was drawn through them. The departures of the individual readings from this curve were then tabulated and subjected to harmonic analysis, which yielded the components of the lunar variations. The nominal accuracy of measurement, which was made every quarter of an hour, was 0·5 km. Though the amplitude of the variation was comparable with the error of each measurement, a strongly marked and statistically reliable semi-diurnal variation was revealed. This could be expressed in the form 0·93 sin $(2t'+112^{\circ})$ km., where t' is the

lunar hour angle. The maximum is therefore attained about $\frac{3}{4}$ hour before the lunar transit. Fig. 84 shows the semi-diurnal variation of the height of the E-layer on the harmonic dial (see Chap. II, Sec. 2*b*). It will be noticed that the points are closely grouped, so that the probable error circle does not enclose the centre of the dial: This shows that the result obtained is statistically significant. Further, they lie not far from the point S, which was derived from an independent series of midday values.

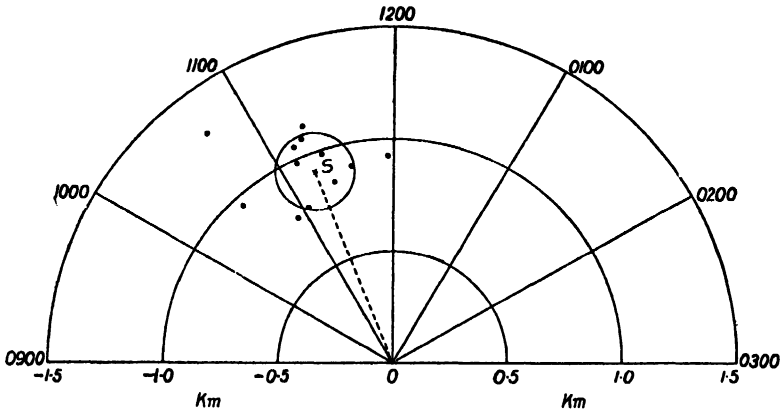


FIG. 84.—Harmonic dial illustrating lunar semi-diurnal oscillation of the height of Region E. (After Appleton and Weekes.)

Observations at other stations (Canberra, Brisbane [259]) have also established semi-diurnal lunar tidal variations of h'_E. It appears, however, that the phase (say the hour of maximum) depends markedly upon the latitude. Thus, while the maximum occurs at Slough (51°N.) about one hour *before* lunar transit, that at Canberra (35°S.) occurs five hours *after* the same. Tidal variations of f°_E (in phase with h'_E) have also been reported [260]; this, however, is not yet well established.

Lunar semi-diurnal tidal oscillations in the D-region, with maximum absorption occurring at 10·9 lunar hours, have been detected by Appleton and Beynon [262].

Lunar tidal variations of the heights and ionization densities of the F_1-layer have also been determined for a number of stations [260]. The amplitudes are of the same order of magnitude as those of the E-layer. The maximum of these variations occurs 6 hours *after* lunar transit. Unlike that of h'_E the phase of $h'F_1$ variation does not appear to show any regular variation with latitude.

For the F_2-region the lunar semi-diurnal tidal variations are much more well-marked than those for the E and F_1 regions. Analysis of only 3-4 years' data is sufficient to reveal the semi-diurnal harmonic [256].

It may be mentioned that a lunar *diurnal* variation of semi-thickness of Region F has been reported by Jones and Jones at College, Alaska, based on 6 months' data [263]. The result, however, needs confirmation.

(ii) *Solar tidal effects.*—Solar tidal variations for the E and F_1 regions have not yet been properly derived. As mentioned earlier, they are difficult to isolate, being superposed on the much larger semi-diurnal harmonic of the solar ionization effect. Further, the amplitudes of the variations are small, being of the same order as the lunar tidal variations. Nevertheless, close examination of the E and F_1 records reveals certain features in their semi-diurnal variations which, according to Martyn, are ascribable to tidal effects.

A simple evidence of the solar tidal motion in the F_2-region is obtained from a study of the times at which h' (or h_{max}) for this region reaches maximum value in course of the day. From the curves in Fig. 85 it is clearly seen that there are two maxima, one at about midday and the other at about midnight. The midday maximum *might* be explained as a heating effect, but it is difficult to explain the midnight maximum as due to the same cause. On the assumption of a tidal motion of 12-hour periodicity both the midday and midnight maxima are explained.

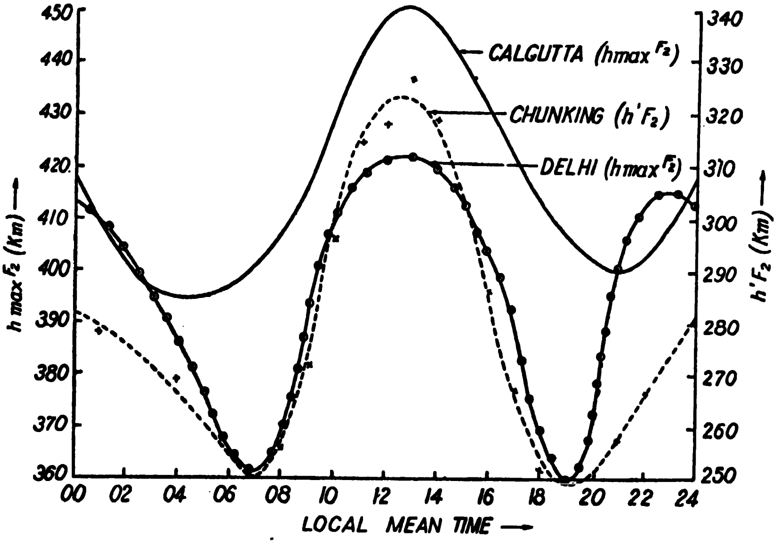


FIG. 85.—Average diurnal height variation for Region F_2 for Calcutta, Delhi and Chungking, 1946–48, illustrating solar semi-diurnal tidal oscillation [260a]. ($h_{max}F_2$ for Calcutta and Delhi, $h'F_2$ for Chungking.)

Monthly mean hourly values of $h'F_2$ (computed from data for a few years) show an average periodicity of almost exactly 12 hours. An interesting point may be noted in this connection. Each maximum shows a seasonal swing, the maximum occurring about one hour before midday (and midnight) in summer and one hour after midday (and midnight) in winter. The fact that in this swing of the maxima from summer to winter, a constant time-difference of 12 hours is preserved supports the hypothesis that the variations observed are due to tidal effects.

Solar tidal effects in the variation of $N_{max}F_2$ have also been observed. This effect, however, is generally masked by the large variation due to the ionizing action of the ultraviolet solar radiation. To observe the effect of this tidal motion it is, therefore, advisable to compare the diurnal variation of N_{max} at two or more latitudes for which the ionizing term (which is proportional to $\cos^{\frac{1}{2}} \chi$) varies little, but for which the tidal term varies markedly. This has been done by Martyn, the three stations chosen for the purpose being Cape York (11° S.), Brisbane (28° S.) and Canberra A.C.T. (35° S.). (All situated near longitude 140° E.) Similar analysis has also been made with records of Calcutta (22·5° N.), Chungking (29·4° N.) and Delhi (28·6° N.). (All situated near longitude 100° E.) In Fig. 86, Curve I depicts the differences between corresponding mean hourly values of $f^\circ F_2$ (which is proportional to $\sqrt{N_{max}}$) at Calcutta and Chungking. Curve II depicts the same quantity for Calcutta and Delhi. The curves clearly show that there is a semi-diurnal variation in $f^\circ F_2$ for pairs of stations.

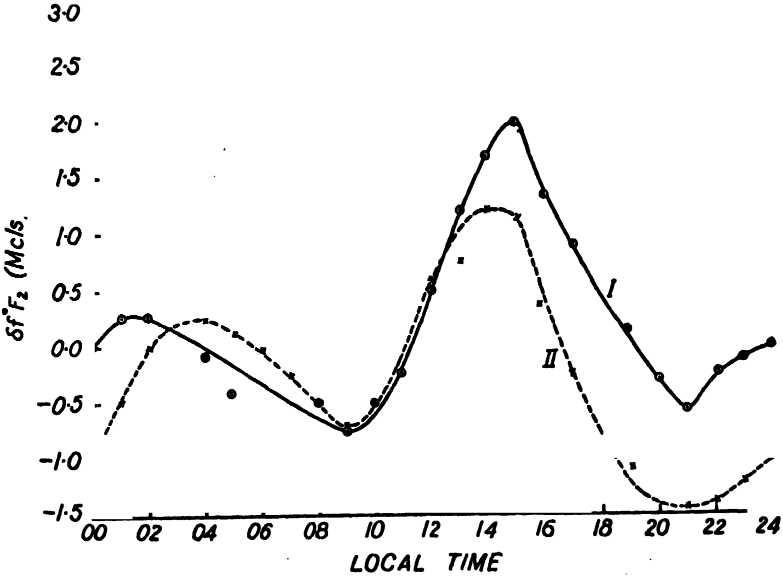


FIG. 86.—Differences between average diurnal variation of $f^\circ F_2$ at Calcutta, Delhi and Chungking illustrating solar semi-diurnal tidal oscillation. Curve I, Calcutta-Chungking; Curve II, Calcutta-Delhi.

(iii) *The electrodynamical theory* (*Martyn*).—The fundamental basis of the so-called electrodynamical theory as developed by Martyn is that though the tidal motions are mainly horizontal, the ions and electrons are constrained to move along the terrestrial magnetic lines of force. This motion has, in general, a vertical component except at the magnetic equator. But even at such low latitudes vertical ion (and electron) drift can occur if a horizontal electric field exists; such a field must occur if the 'dynamo theory' of magnetic variations is valid (see Chap. VII, Sec. 5). It has been shown by Martyn that these characteristic motions of ions and electrons,

besides explaining the observed tidal variations, also explain many of the known anomalies of the behaviour of F_2-layer, and the small but significant departures of the behaviour of E and F_1 regions from that of the ideal Chapman region. Another fact that has emerged from these considerations is that the usual methods of measuring the recombination coefficient α from observations of the diurnal variation of ionization (see Sec. 12) need reconsideration.

As indicated above, the vertical drift velocity (v) of the ions (and electrons) consists of two parts. The first (v_1) is due to the vertical component of the air motion along the magnetic field (the existence of which is deduced from simple consideration of the electromagnetic forces) and the second (v_2) is due to the polarization field developed by the motions of the ions across the magnetic field (to the existence of which Martyn first directed attention).

The origin of v_1 is as follows. It can be shown that, in the presence of the terrestrial magnetic field, the drifts of the ions due to the forces developed by the tidal air motions will have the components (for example, see *Geomagnetism* by Chapman and Bartels, pp. 531–535; also Chap. VII, Sec. 5 of this volume),

$$\dot{x}' = c_{\perp}\frac{\nu^2}{\nu^2+\omega^2},$$

$$\dot{y}' = -c_{\perp}\frac{\nu\omega}{\nu^2+\omega^2},$$

$$\dot{z}' = c_{\parallel},$$

where $c_{\parallel}$ and $c_{\perp}$ are components along z and x axes respectively of the horizontal tidal air velocity, $\omega = He/m$ and ν frequency of collision of the ions with the neutral particles. (The axes are so chosen that z lies along the magnetic field and x along the component of the tidal force transverse to the field.)

There is thus a vertical drift velocity as all the three component velocities have component in the vertical direction. However, of these, the $\dot{y}'$-component depends on the sign of the ionic charge. The $\dot{x}'$ and $\dot{z}'$-components are independent of the sign and, of these, the former is negligible if (as in the case of the F_2-region) v is much smaller than ω. Hence, under such condition v_1 is nearly equal to $c_{\parallel}$ sin I (I, magnetic dip), being thus very nearly the vertical component of the air-velocity along the magnetic field.

The polarization effect, to which v_2 is due, is developed by the velocity component $\dot{y}'$ depending upon, as already mentioned, the ionic sign. As a result of this ions of opposite sign move in opposite directions giving rise to an electric current and thus to a polarization of the medium. The component of any resulting polarization field perpendicular to the magnetic field ($E_{\perp}$) will produce ionic drift. In particular, vertical ionic drift will be produced by the component $E_{\perp}'$ of $E_{\perp}$ lying in the horizontal plane

containing H. The magnitude of this vertical ionic drift (v_2) is shown by Martyn to be:

$$v_2 = \frac{E_\perp'}{H} \cdot \frac{\omega^2}{\nu^2+\omega^2} \cos I.$$

Further, the sense of this drift is opposite to that of the previous one.

At any latitude the total vertical ionic drift (v) is thus the sum of the two drift velocities: $v = v_1 + v_2$.

Both v_1 and v_2 necessarily vary with latitude. In middle latitudes the air velocity along the magnetic field has a large vertical component and v_1 has a maximum. In low latitudes the magnetic field is nearly horizontal and the value of v_1 is, therefore, small. On the other hand v_2 has its maximum at the magnetic equator, where $\cos I$ is greatest. Both v_1 and v_2 tend to zero at the poles, the former because a horizontal wind has no component along the (vertical) magnetic field, the latter because $I = 90°$. Thus at all places save the poles there are vertical ionic drifts, though for causes distinctly different in high and low latitudes. And, since v_1 and v_2 are opposite in sign, the vertical drift in high latitudes will be in opposite direction to that in low latitudes. The reversal of phase occurs in an intermediate latitude where v_1 and v_2 are numerically equal. This latitude is found to be about 35°. It is to be noted that the focus of the dynamo current system which produces the quiet day magnetic variation also lies at about the same latitude (see Figs. 12 and 13, Chap. VII, Sec. 3(*c*)).

The drift velocity also varies with time. At any point it has a semi-diurnal 12-hour period and a seasonal annual period. (This may be expected from the fact that the horizontal solar tidal air motion has predominant semi-diurnal and seasonal components.) Further, in the F-region and above, the amplitude of v decreases with height on account of damping due to viscosity and thermometric conductivity with height (see Fig. 3, Chap. I). In view of the above the expression for v may be assumed to be of the form $v = v_0 e^{-\gamma z} \sin(\omega t + \sigma)$ where γ is of the order unity and ω ($= 360°/12$) is 30° per hour. (The initial phase σ is also sometimes assumed to change with height.) The magnitude of v_0 in the F-region is found to be of the order 300 cm./s.

Now, it is easy to see that other factors remaining constant, if the vertical drift velocity decreases with height, i.e. if there is a strong velocity gradient, the ionic concentration at any point will change with time. This fact introduces a new term

$$\left[\frac{\partial}{\partial z}(Nv)\right]$$

in the expression for the time rate of change of the ionic concentration (see Eq. 156). This, according to Martyn, is

$$\frac{\partial N}{\partial t} = I(z, t) - \alpha N^2 + \frac{\partial}{\partial z}(Nv), \quad \ldots \quad \ldots \quad \ldots (156)$$

where I—rate of ion production per cm.^{3},
Z—reduced height (h/H, h being measured from a datum level),
N—density of ionization,
v—drift velocity measured positively downwards.

This is the fundamental equation of the electrodynamical theory, showing how the normal Chapman distribution with height and the normal Chapman variation with time will be affected by the drift velocity developed by the tidal action. The extent to which the drift velocity will affect the normal distribution will, of course, depend on the magnitude of the third, the 'drift' term $\frac{\partial}{\partial z}(Nv)$ compared to the first two in the equation.

For the F_2-region, when v is directed upwards, the 'drift' term becomes the most important term. This is because, the velocity gradient is high on account of increase of viscosity with height and, also, in the second term, α is known to be small—some two magnitudes smaller than α for the E and F_1 regions. Further, the rate of ion production (first term, I) is also small. Hence, for such case for the F_2-region the continuity equation becomes approximately equal to $\frac{\partial N}{\partial t} = \frac{\partial}{\partial z}(Nv)$. This equation becomes specially true for the night-time condition when $I = 0$, and has been solved for different assumed variations of v with height and time [256, 257, 264, 260]. It is found that for a low latitude station the distortion of an initial Chapman region with the passage of time as calculated from this equation, generally agrees with the behaviour of the F-region at night after 21 hr., i.e. from the time the drift velocity becomes directed upwards.

In contrast with the above, for the E and F_1 regions (as also for the F_2-region when v is directed downward) all the three terms in the equation are of importance. Hence, to compare theoretical values with the observed ones, one has to solve the equation taking account of all the three terms. The solution of the complete equation is necessarily difficult, but has been made by Kirkpatrick [265] for the E-region and by A. P. Mitra [266] for the night time F_2-region.

According to Kirkpatrick's calculations the departures of the values of the E-region parameters from the Chapman values are, as expected, small and of the same order as observed. In particular, the height perturbation is in phase with the drift, i.e. in opposite senses in the latitudes north and south of 35° and is thus in excellent agreement with the observed results on the lunar $\delta h'E$ variation in high and low latitudes [e.g. between Canberra and Brisbane (low latitude) on the one hand and Slough (high latitude) on the other].

According to A. P. Mitra's calculations, the effect of the drift velocity on the night time variation of the F_2-region is independent of recombination effect and adds to or subtracts from the latter according as the drift velocity is downwards or upwards.

Analysis of ionospheric data also enables one, with the help of Eq. 156, to determine the amplitudes and phases of the tidal drifts as also the value of the effective recombination coefficient corrected for the effect of the same. For example, according to A. P. Mitra's calculations [266] the seasonal and semi-diurnal tidal drifts (solar) for Calcutta are found to be of nearly equal magnitude as demanded by the electrodynamical theory [248]. The resultant drift velocity has been found to be 12 km./hr. in both summer and winter and 17 km./hr. in the equinoxes, being maximum downwards at 1630 in summer, at 1500 in the equinoxes and at 1330 in winter. The values of the recombination coefficient calculated by him are all of the order of 10^{-11} cm.^{3}/s., the winter value (3×10^{-11} cm.^{3}/s.) being twice as large as the summer value ($1{\cdot}5\times10^{-11}$ cm.^{3}/s.). The seasonal variation of α for the F_2-regions, as observed by many workers, is thus due partly to temperature and pressure variations, and partly to tidal effects.

It may also be mentioned that two hitherto inexplicable characteristics of F_2-region variation in summer, namely, decreased day-time N_{max} and marked bifurcation of F_1 and F_2 in low latitudes can be explained by means of the electrodynamical theory. The first is attributed to the downward transportation of ions in summer to regions where the recombination rate is very high, and the second to the steep gradient of drift velocity (and/or recombination coefficient) in the upper part of the combined F-layer.

From the brief account of the ionospheric tidal theory given above it is clear that the theory is a definite step forward in our understanding of tidal phenomena and anomalous variations in the ionosphere. The theory, however, is still in the formative stage. Much more work, both observational and theoretical, has to be done, to fill the many gaps that still exist.

(*d*) Travelling disturbances

In subsection (*b*) we have discussed winds in the E-region in which ionospheric matter (a 'patch' or 'cloud' of intense ionization) appears to be transported horizontally from over one place to another. In contrast to this, observations in the F-region shows that there are occasions when some characteristic feature of ionization, e.g. a characteristic hump in the h'—t curve (and not ionospheric matter), appears to travel as a disturbance wave from over one observing station to another [267, 268]. A typical disturbance like this travels with a speed of 10 km./min.; its wavelength is some hundred of kilometers and the apparent period of the order of half an hour. Martyn [269] has examined the origin of such travelling disturbances in the F_2-region and ascribes them to horizontally travelling 'cellular' waves of the type first investigated by Lamb. (Martyn also ascribes the microbarometric oscillations in the troposphere—known for a long time—to such waves.)

In a cellular wave the air particles move in vertical plane in elliptic paths bounded between two horizontal surfaces. The bounding may

occur as a result of sudden increase in either the lapse rate or the temperature, or, a high gradient of the wind velocity (which latter may produce the disturbance leading to the formation of cellular wave). In the ionosphere, the 80 km. level where there is a high lapse rate might be the lower boundary. The upper boundary is, however, difficult to locate. A high lapse rate is found to be necessary even if there is a high wind gradient, and, a high lapse rate is not possible in this region. Martyn, therefore, proposes as an alternative a low value of the specific heats ratio (1·0 instead of 1·4), possibly as a result of the dissociation of N_2. (See Chap. V, Sec. 3.)

The motion of the cellular waves explains a number of apparently uncorrelated and otherwise inexplicable F_2-region phenomena. These are:

(1) Travelling disturbances, or, a rhythmic change of height (of a layer of constant ionization density) of the F_2-layer and of its electron concentration as mentioned at the beginning [267, 268].

(2) Ionospheric records sometimes show as if reflecting electron clouds are coming down vertically into the F_2-region [270].

(3) Systematic errors in D.F. measurements for propagation *via* F_2-region. Observations on frequencies 6–15 Mc./s. and for transmission distance 90 to 5,000 km. show that the root-mean-square-error (standard deviation) in the bearing measurement may, for the shorter distance, be as much as 8° [271, 272, 273].

All these phenomena can be explained if it is recalled that the electrons (and ions), while participating in the cellular wave motion, are constrained to move along the lines of force. (The elliptic paths of the particles are of rather high eccentricity with the major axis horizontal). A little reflection shows that if the lines of force have a large vertical component, then, as the electrons cannot move across the lines of force, in one part of the elliptic path the electrons will move downwards and in another part upwards, both along the lines of force. Surfaces of equal ionization will therefore be continually bending, as if subject to a wave passing over it.

The travelling disturbances as mentioned in (1) are thus simply the passage of these ionization inequalities over the observing station with the velocity of the cellular waves.

The phase of the cellular oscillatory motion has a height gradient. This gradient makes horizontally travelling disturbances appear to have a vertical component of motion thus simulating vertically moving electron clouds. (See, for example Fig. 6, ref. [269].) This explains phenomenon (2).

The 'tilts' or 'slopes' in the surfaces of constant ionization produced by the travelling disturbances cause deviation of the rays reflected from them from the great circle path and account for phenomenon (3). (See Fig. 6, *loc. cit.*)

15. LUXEMBOURG EFFECT—CROSS-MODULATION

(i) *Introduction.*—The 'Luxembourg effect' or ionospheric cross-modulation may be described as the modulation of the signal from a

powerful station impressing itself upon the carrier wave from another station operating on a different frequency. The effect is caused by the interaction of the radio waves *via* the ionosphere. When this happens it is found that a receiver tuned to the second—the wanted—station has, in its background, the programme (modulation) of the first—the disturbing station. The phenomenon is illustrated in Fig. 87.

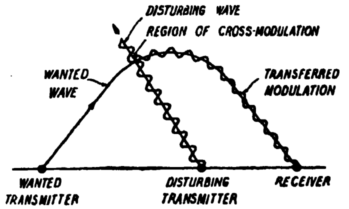


FIG. 87.—Illustrating 'Luxembourg effect' or the phenomenon of ionospheric cross-modulation. The modulation of the wave from the disturbing transmitter is transferred *via* ionosphere to the wave from the wanted transmitter. (After, Huxley and Ratcliffe.)

The existence of cross-modulation was first reported by Butt in 1933 (see ref. [274]). It was found that in the reception of the Beromünster programme at Eindhoven, there was often programme from Luxembourg in the background. This was followed by studies by Tellegen [275], van der Pol and van der Mark and others [276, 277, 278].

A detailed theory of the phenomenon has been worked out by Bailey and Martyn [274] by utilizing the results of investigations of Townsend and Tizard [279] on the motions of electrons in air at low pressure and under the influence of electric forces. In what follows we will first give, without going into mathematical details, an account of the physical processes involved in the phenomenon in the words of Huxley and Ratcliffe [280].

(ii) *Theory—the physical processes involved* (*after Huxley and Ratcliffe*).—'The absorption of a wave passing through the ionosphere occurs in the following way. In the absence of a wave the electrons in the ionosphere are in thermal equilibrium with the gas molecules, so that their mean energies have the same magnitude Q_0. An electron loses energy in some collisions with gas molecules and gains energy in others, but on the whole the gains and losses neutralize each other.

'When, however, a radio wave traverses the medium an electron acquires extra energy from the wave along each free path and at first this extra energy is not, on the average, entirely lost in the collision at the end of the free path. The mean agitational energy Q of the electron therefore begins to increase to a value greater than the thermal energy Q_0. But when Q exceeds Q_0 the statistical energy balance is upset and, on the average, energy is transferred from the electron to the molecules at each collision. When $(Q-Q_0)$ is small compared with Q_0 we may assume that the average energy lost by the electron in each collision is equal to $G\ (Q-Q_0)$ where G is a constant which has been determined by means of laboratory experiments. In the presence of the wave, therefore, the mean agitational energy Q of an electron continues to increase until the mean energy $G\ (Q-Q_0)$ lost by it at each collision is equal to the mean work done on it by the radio wave in the free path between two collisions.

'It is of interest to consider the magnitudes of the quantities just mentioned. Let us take the case of a 100-kW sender of frequency 200 kc./s. acting like a Hertzian doublet and radiating at an elevation of 45° a wave which is absorbed at a level of 85 km. (radiation from Droitwich at 45°). Then we have:

(i) G for air* is found experimentally to be equal to $1{\cdot}3\times10^{-3}$.

(ii) The mean thermal energy Q_0 of an electron at 300°K. is $6{\cdot}2\times10^{-14}$ erg.

(iii) The mean oscillatory energy communicated from the wave to the electron in each free path (P_1/ν) is $5{\cdot}5\times10^{-17}$ erg.

'We have already seen that this last quantity must be equal to $G(Q-Q_0)$ so that we find $Q-Q_0 \fallingdotseq 4{\cdot}2\times10^{-14}$ erg. which is seen to be $0{\cdot}67\,Q_0$. We therefore conclude that the wave heats up the electrons until their mean energy exceeds the thermal energy Q_0 by 67%. Their temperature is thus increased 67% and the frequency of their collision with molecules is increased 33%.

'Another wave traversing this part of the ionosphere is also absorbed, so that its field E decays according to the law $E = E_0 \exp(-\kappa x)$ and the absorption index κ is proportional to the frequency of collision between electrons and molecules. The absorption of this (the wanted) wave is therefore increased when the electrons are heated up by the presence of the disturbing wave. It has been shown by direct experiment that the absorption of a wanted wave is increased when a strong unmodulated disturbing wave is switched on.

'The energy lost by an electron to a molecule is proportional to its excess thermal energy $(Q-Q_0)$ so that, if the electrons are all heated up and then left to themselves, they will cool to the temperature of the molecules in an exponential manner with a relaxation time τ depending on the product $G\nu$ where ν is the frequency of collision. If the amplitude of the wanted wave could be observed rapidly enough it would be found to return to normal with a time-constant τ after the removal of the disturbing wave.

'Now suppose that the disturbing wave is modulated sinusoidally at the angular frequency ω. There will result an alternate "heating" and "cooling" of the electrons, and hence a periodic variation of the absorption of the wanted wave, which will show itself on reception as a modulation of this, originally unmodulated, wave. The magnitude and phase of the transferred modulation depend on the modulation frequency ω because the ability of the electrons to "heat" and "cool" sufficiently rapidly depends on the relation between the frequency and the natural relaxation time τ.

'For modulation frequencies much less than $(2\pi\tau)^{-1}$ the heating and cooling is large and is in phase with the original modulation; for frequencies

---

* G has been measured in the laboratory for air only; one is not likely to make a significant error by assuming this value to apply in the ionosphere.

much greater it is small, and lags $\pi/2$ in phase: the behaviour is like that of any other "relaxation process" such as is encountered in Debye's theory of dielectrics containing polar molecules or in the application of a sinusoidal e.m.f. to a combination of capacitance and resistance.

'Simple considerations of the kind given here lead directly to the following deductions:—

(*a*) Cross-modulation occurs only on the ionospheric wave and not on the ground wave, and hence is noticeable only when the ionospheric wave is much stronger than the ground wave.

(*b*) Cross-modulation is most pronounced when the wanted wave is strongly absorbed in the region where the disturbing wave is itself absorbed. For some combinations of frequencies the two absorbing regions will not overlap. For some wanted frequencies, especially the higher frequencies at night, the absorption may be too small. In both these cases cross-modulation is small.

(*c*) The energy supplied to the electron in each free path by the disturbing wave is proportional to the power in that wave and to the square of the wavelength. Transferred modulation is therefore most noticeable when the disturbing station has a high power and a long wavelength. It is generally noticed, in practice, with the long-wave broadcasting stations, such as Luxembourg. The fact that it has not been observed on still longer wavelengths must be ascribed to the non-overlap of the appropriate absorbing regions mentioned at (*b*).

(*d*) The proportionality between the coefficient of transferred modulation and the *power* of the disturbing station implies that the transferred modulation will contain an octave component of the modulation frequency, but not higher harmonics. This has been observed and measured.'

(iii) *Experimental observations.*—Perhaps the most important experimental verification of the mathematical theory of cross-modulation (not given here) is that of the law of the dependence of cross-modulation (coefficient of transferred modulation) on the modulation frequency f of the disturbing wave. The law is

$$T_f = T_0 \frac{1}{\sqrt{1+\left(\frac{2\pi f}{G\nu}\right)^2}}, \quad \ldots \quad \ldots \quad \ldots \quad (157)$$

where T_0 represents the value to which T_f tends as the impressed modulation frequency approaches zero ($T_0 = \beta M P_1/G\bar{Q}$). Here, the symbols are:

f—modulation frequency of the disturbing wave,

T_f—coefficient of transferred modulation,

ν—collisional frequency of electron with molecules,

G—a constant (*vide supra*),

β—is so defined that the wanted wave absorption $\int \kappa_w ds$ expressed in nepers is equal to $\beta\nu$,

M—coefficient of modulation of the disturbing (unwanted) wave,

P_1—power supplied to a single electron by the unmodulated wave,
$\bar{Q}$—mean value of the agitational energy of the electron in the presence of the modulated wave.

Experimental observations show that this relation holds up to modulation frequencies of 800 c/s [280, 281, 282, 283, 284]. It appears that at higher frequencies the observed values of T_f are less than the calculated value [280].

The value of G may be deduced experimentally and has been found to lie in the range 160 and 320 sec.$^{-1}$ [99]. Assuming $G\nu/2\pi$ to be 240 sec.$^{-1}$ we can write

$$T_f = T_0 \frac{1}{\sqrt{1+\left(\frac{f}{240}\right)^2}} \quad \ldots \quad \ldots \quad \ldots (158)$$

Fig. 88 shows how T_f varies with the modulation frequency f according to the above relation.

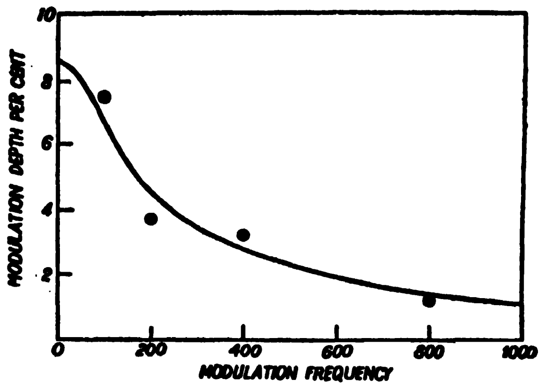


FIG. 88. Illustrating how the depth of the transferred modulation varies with the modulation frequency of the disturbing wave. The dots plotted are for the wanted carrier wave of the Beromünster station. (After van der Pol and van der Mark.)

It has also been observed that as demanded by the theory, the coefficient of the transferred modulation is proportional to the power of the disturbing transmission. (Observations have been made up to a power of 560 kW. [280, 283].)

The presence of an octave component of the modulation frequency, which follows as a consequence of the above, has also been observed and measured [284].

It has been found that the coefficient of the transferred modulation remains proportional to the depth of modulation of the disturbing wave even when the latter is 80% and the transmitter power is 560 kW. [280].

It has been shown that the phenomenon of cross-modulation is most marked when the disturbing transmitter is situated within 100–200 km. from the midpoint of the transmission path of the wanted wave.

According to experiments of Ratcliffe and Shaw [283] the greatest transferred modulation is produced when the wanted wave is in the medium wave band (in the range 800–1000 kc./s.) and comes from a sender distant 300 or 400 km. and the disturbing wave is in the long wave band (200 kc./s. and 167 kc./s. had been used). These results are for observations at night.

According to theory the cross-modulation effect is to increase with the lowering of the frequency of the disturbing wave. It appears from the existing observational results that this holds till the frequency 167 kc./s. is reached. The effect, however, appears to decrease with further reduction of frequency [283]. The highest frequency of the disturbing wave, up to which cross-modulation has been observed is in the neighbourhood of gyro-frequency.

Ratcliffe and Shaw have suggested an interesting method of measuring the height at which interaction takes place from an examination of the phase of the modulation impressed [283]. In the experiments carried out most of the measured heights were found to lie within 82 and 85 km. The height at which the interaction takes place has also been estimated by Huxley [99*a*] from the measured value of $G\nu$ (*vide infra*). The value found is 83 km. and is thus in good agreement with that obtained by the 'phase method'.

(iv) *Measurement of collisional frequency.*—From Eq. 157 it is evident that if T_f is known for two different modulation frequencies of the disturbing station, then the value of the quantity $G\nu$ can be obtained. Further, if G is determined experimentally, then ν, the collisional frequency of the electron in the region of the interaction, may be determined. Van der Mark and van der Pol were the first to determine ν from such cross-modulation observations [276]. They measured the modulation (T_f) impressed upon the Beromünster carrier for different values of the modulation frequency and for a fixed value of the coefficient of modulation. From a plot (Fig. 88) of the observed results (modulation frequency as a function of the percentage depth of the impressed modulation) they obtained T_f as 4·8 per cent and 1·8 per cent for the modulation frequencies 200 and 800 c./s. respectively. Substituting these values in Eq. 157 and taking the ratio we get,

$$\left(\frac{1\cdot 3}{4\cdot 8}\right)^2 = \frac{(2\pi\times 200)^2+(2\cdot 6\times 10^{-3})^2\nu^2}{(2\pi\times 800)^2+(2\cdot 6\times 10^{-3})^2\nu^2}.$$

ν is thus found to be $2\cdot 1\times 10^6$/s. (van der Pol and van der Mark took the value of $G = 2\cdot 6\times 10^{-3}$ as deduced from experiments of Townsend and Tizard [279]). ν may also be determined by a slightly different procedure after Huxley, Foster and Newton [284]. Eq. 157 may be written as

$$f^2 = \frac{T_0^2}{T_f^2}\frac{G^2\nu^2}{4\pi^2} - \frac{G^2\nu^2}{4\pi^2},$$

showing that the curve representing f^2 as a function of $1/T_f^2$ is a straight line. Hence $G\nu/2\pi$ may be obtained from the intercept on the f^2 axis

of the straight line drawn through the experimentally observed points. From observations on the transmission from Lisnagarvey (1050 kc./s.) with Droitwich (200 kc./s.) as the disturbing station the above named authors obtained the value $G\nu$ lying between 1140 and 2860. Using the experimentally determined result of G by Huxley and Zaazou [99] as $1{\cdot}3\times10^{-3}$, the value of the collisional frequency ν was found to be $1{\cdot}2\times10^{6}$/s. Shaw [290] obtained the average night-time value to be $1{\cdot}4\times10^{6}$/s. at 92 km.

(v) *Effect of the terrestrial magnetic field.*—In the above discussion we have not taken into account the effect of the terrestrial magnetic field. The inclusion of this effect necessarily makes the analysis complicated. However, some of the salient points of the effect can be easily seen from simple consideration of the magneto-ionic formula.

We have seen that the primary cause of cross-modulation is the increase of the mean agitational energy of the electrons by absorption of energy from the disturbing wave. The magnetic field may profoundly modify this absorption, depending on the wave frequency and the direction of propagation. We have to consider the cases of the ordinary and the extraordinary waves and, for each of these, the absorption will depend on the mode of propagation—whether it is quasi-longitudinal or quasi-transverse. (See Sec. 5). It is easily shown that the extraordinary wave will be strongly absorbed near the gyro-frequency for all directions of propagation. (This phenomenon has been specially discussed by Bailey [281, 285, 274] and is called by him *gyro-interaction*. See below.) Further, for the ordinary wave, for the quasi-transverse mode of propagation, the absorption, and hence the cross-modulation, should increase steadily with the lowering of the frequency of the disturbing wave. The cross-modulation produced by low frequency wave is, therefore, expected to depend on the direction of propagation.

Of the above conclusions the phenomenon of gyro-interaction or gyro-resonance effect has excited considerable interest. However, there is doubt about experimental evidence of the phenomenon. Thus, while according to Bailey [286] and also to Cutolo [287, 288, 289] gyro-resonance effect is observable under certain conditions, according to experiments of Shaw [290], the predicted enhancement of cross-modulation does not exist.

In conclusion, it may be mentioned that the phenomenon of ionospheric cross-modulation is of interest for more than one reason. It has provided us, for the first time, with a means of making controlled alterations in the ionosphere. Further, it provides a link between laboratory and ionospheric studies of the molecular and electronic parameters of the upper atmosphere.

16. WEATHER AND THE IONOSPHERE

Relationships between barometric and weather conditions in the troposphere and ionospheric conditions in the upper atmosphere have been reported from time to time from different parts of the world. It

is not easy to foresee how such relationships could occur. We give below a few instances of the same without commenting upon their possible causes.

It has been observed that variations of minimum height of the F-region and of the average E-ionization tend to follow the variation of barometric height [292]. Similar correlation (on an annual basis) between the lowest virtual height of E-region (on 100 kc./s.) and ground temperature has been reported from Stanford, California, U.S.A. [192]. It was found that the height tended to be lower in winter than in summer as did the temperature. Some slight correlation between pressure at 13 km. and critical frequencies of E and F_2 regions has been reported [292]. Two other associations, rather striking in nature, between weather phenomena and variations of ionospheric characteristics have also been reported, one from Australia (Sydney) [293, 294] and the other from China (Shanghai) [291]. These are briefly described below.

The weather in Australia according to meteorological observations is controlled by a fairly regular succession of anticyclones passing from west to east. In the region between two successive anticyclones there is a zone in which the air characteristics change discontinuously. To the east of this zone warm, water vapour laden air flows from lower latitudes; to the west cool and dry air flows from high latitudes. The cold and warm masses ultimately develop into cyclonic depressions. Now, the two ionospheric stations, Mount Stromlo (35° 15′ S., 149° 10′ E.) and Liverpool (33° 55′ S., 150° 59′ E.) are situated near the path of the centre of the tracks of the anticyclones. In course of observations extended over a period of nineteen months, carried out at these two stations it was found that on the days on which the region between lat. 29° S. and lat. 36° S., and long. 140° E. and long. 156° E. was free from 'frontal conditions,' Region F_2-ionization was higher, on the average by 6 per cent near Sydney (Liverpool) and by about 11 per cent near Canberra (Mount Stromlo) than on other days. Further, it was noticed that while the values of ionization density at the two observing stations (separated only by a distance of 250 km.) were usually equal, there was a strong tendency for the ionization at Canberra to be lower than that at Sydney on 'frontal' days. (It is to be mentioned that the days on which the F_2-region ionization was disturbed due to magnetic storms, were discarded from the analysis.)

The other correlation that has been observed at the Zi-Ka-Wei observatory, Shanghai is of a similar nature [291]. It was found that when observations are made, any day of the year, after sunset or before sunrise, with the ionosphere apparatus set to the mean E-critical frequency of the place (6 Mc./s.), the following relationships hold between the occurrences of the E, F_1 and F_2 echoes on the one hand, and the future movement or behaviour, on the other, of the three main air masses—the polar, the maritime and the equatorial—which cause weather all over the world. If an E-echo is obtained then the maritime (Pacific) air mass would come over the place or, if it were already there, it would stay over the place. When an F-echo is observed, then the polar (Siberian) air

mass would make the weather. If an F_2-echo is detected, it means that tropical air is arriving, or if it is already there, it will continue to stay. If, when an E-echo is obtained, there be a typhoon on the weather map some 200 miles away, then it could be predicted that the maritime air mass would bring the cyclone dangerously close to the station. If, on the other hand, an F-echo is obtained then, it would mean that there is no danger to the station as the typhoon would recurve. According to the author the relationships are so persistent that (since the air masses are responsible for the weather) it has been possible to make forecasts, with exceptional success, on the future weather characteristics—'dry or damp, overcast or clear, windy or calm, good or bad visibility, high or low temperatures, etc.'

It is difficult to give any theoretical explanation of these curious associations. One has to assume that the atmospheric constituents in the ionospheric regions must be changing from day to day—and that rapidly—in order to maintain association with meteorological conditions near the ground. The whole subject demands further study and observation in other parts of the world.

CHAPTER VII

ELECTRICAL CURRENTS IN THE UPPER ATMOSPHERE: TERRESTRIAL MAGNETIC VARIATIONS

1. INTRODUCTION

In this chapter we shall discuss the morphology and origin of the electrical current systems in the upper atmosphere (and, also a part, in the extra-terrestrial space) the existence of which is revealed by the study of transient variations of the terrestrial magnetic elements, e.g. the horizontal and vertical intensities and declination.

(a) Transient terrestrial magnetic variations and upper atmospheric current systems

The terrestrial magnetic field as measured at the surface of the earth is not constant in time. Its intensity varies and the variations, according to their rates, may be broadly classified under two heads: *secular variation* and *transient variation.* The former produces perceptible effects in periods measured in decades or centuries and the latter—more rapid ones—have periods measured in days, hours or even minutes. The origin of the secular variation is still ill understood. The transient variations are now known to be due to world-wide current systems, partly in the conducting regions of the upper atmosphere and partly in extra-terrestrial space. The amplitudes of the transient variations are always very small compared to the total intensity of the field, amounting only to a few thousandths of it. These variations are measured in terms of a small unit gamma (γ), equal to 10^{-5} gauss. But though small, the variations have certain characteristic and persistent features which, even a casual inspection of the hourly records of magnetic data kept at an observatory seldom fails to reveal. Fig. 1 depicts, for instance, a record of the east component (Y) of the magnetic field for a number of successive days which were all magnetically quiet. The regular daily rise and fall is strongly reminiscent of the diurnal variation of barometric pressure as depicted in

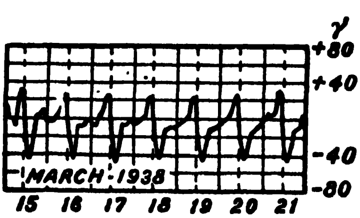


Fig. 1. Record of hourly means of the east component (Y) of the magnetic field at Niemegk Observatory (near Potsdam) on successive magnetically quiet days.

Fig. 1, Chapter II. Indeed, as we shall see later there are strong reasons to believe that the two are causally related. It is obvious from the figure that the period of the regular variation is related to the solar day. Careful

analysis shows that superposed on these regular variations there are also variations, much feebler but none the less as regular, the period of which is related to the lunar day. Occasionally, the regular variations are masked by disturbances. Closer study shows that these disturbances have certain regular features. The task of separating the various types of variations and disturbances from one another is necessarily difficult and laborious. Magnetic data from a large number of observatories, recorded over many years, have to be collected for the purpose. And, the computation is made all the more difficult by the fact that the world distribution of the observatories—there are some eighty of them—is not at all that could be desired; there are too many in the European countries and too few in the southern hemisphere.

In the present chapter we shall make detailed study of the three types of variation mentioned above, namely, (*a*) those related to the solar day, designated S, (*b*) those related to the lunar day, designated L and (*c*) disturbances, designated D.

Our interest in this study lies in the fact that the immediate cause of these variations (except of those known as *sudden commencement* magnetic storms) is the generation of world-wide electric current systems in the ionized regions of the upper atmosphere. In anticipation of what follows we may mention here that for the case of the quiet day type of variations (as in Fig. 1) related to the solar or to the lunar day, the current systems are produced by tidal motions in the upper atmospheric regions as discussed in Chapter II. For the variations associated with the magnetically disturbed days, the corresponding current systems are produced presumably by the entry into the terrestrial atmosphere of high velocity neutral streams of corpuscles (i.e., containing positives and negatives in equal number) of solar origin. It is remarkable that the geomagneticians have not only been able to separate, by laborious analysis, the different types of variation from the mass of accumulated data, but have also been able to construct the probable upper atmospheric current systems which would produce these variations.

It should be mentioned that it was Balfour Stewart who first suggested that the most probable location of the current systems is in the conducting upper regions of the atmosphere. His main argument was that while no noticeable change is observed either in the lower atmosphere or in the body of the earth in course of a solar cycle, marked change in the amplitude is a characteristic of the S_q variation during the solar cycle. There are also current systems within the body of the earth [Sec. 6(*g*)]. But, their magnitude is only about one-third of the upper atmospheric system and, as will be seen later, there is ample evidence to show that the earth-currents are merely currents induced by the upper atmospheric current systems.

(*b*) The earth as a magnet

The earth, so far as its magnetic field at large distances is concerned, behaves to a very close approximation, like a uniformly magnetized

sphere. For such a sphere of radius a, the intensity of magnetization J, the surface value of the horizontal intensity H_0 at the (magnetic) equator and the magnetic moment M are related by

$$M = \frac{4}{3}\pi a^3 J = a^3 H_0. \qquad \text{.. .. ..} \quad (1)$$

The magnetic moment of the sphere is the moment of the equivalent dipole. By choosing the most appropriate value of J (or M) and of the direction of magnetization one can obtain a magnetic field which most closely represents the earth's field. The value of J (or M) and of the direction of magnetization which give this closest fit may be determined by the method of spherical harmonic analysis. The part of the earth's field which can be represented in this way is called the *regular field* of the earth. The remaining part of the field is called the *irregular field.* It has been found by analysis that the regular part of the field is the same as that due to a uniform magnetization of 0·075 C.G.S. unit per cm.^{3} throughout the entire volume of the earth or that due to a magnetic moment of $8{\cdot}19 \times 10^{25}$ C.G.S. units. The magnetic axis, i.e., the diameter along the direction of magnetization meets the surface of the earth at points called the magnetic poles. The approximate positions of these *regular* magnetic poles or *axis*-poles are [1]:

North Magnetic Pole—Lat. 78·6° N.; Long. 70·1° W.
South Magnetic Pole—Lat. 78·6° S.; Long. 250·1° W.

The magnetic axis thus does not coincide with the axis of rotation but is nclined to it at an angle of about 11·5°.

It is to be noted that the magnetic *dip*-poles, i.e., the points on the surface of the earth where the magnetic needle stands vertically ($H = 0$ and $I = 90°$) are different from the axis poles or the regular magnetic poles as defined above. The positions of the dip-poles depend upon local conditions. They are not quite antipodal; according to Vestine they are located approximately as follows in the 1945 epoch:

North dip-pole .. 72·8° N., 262° E.,
South dip-pole .. 68·2° S., 145° E.

If the earth had only its regular field, the lines of equal magnetic force would be circles centred on the magnetic axis, i.e., parallel to the magnetic equator. Actually this is not so; the permanent part of the irregular field having its origin inside the earth distorts their circular form. The intensity of the irregular field, however, decreases much more rapidly than that of the regular field so that at great distances the terrestrial field may be regarded as regular except insofar as it is modified by currents in the upper atmosphere or, in extra-terrestrial space, to which we shall presently refer. It is interesting to note that the intensity of the regular part of the terrestrial magnetic field decreases inversely as the cube of the distance from the earth's centre while that of the gravitational field falls off inversely as the square of the distance.

Though by far the largest part of the earth's field is found to have its origin inside the earth, careful analysis by Vestine [1] shows that a small but definite proportion—less than one per cent—of the field is of external origin. This latter consists of two parts; one possessing a potential and the other a non-potential field. The existence of the non-potential field implies that there is flow of electric current from air to earth or *vice versa*. Owing to the presence of this current the line-integral of the horizontal magnetic force round a closed curve on the surface of the earth does not vanish. The existence of earth-air currents is well known to students of atmospheric electricity. The magnitude of this observed current, however, is wholly inadequate to explain the non-potential field. One may, for instance, calculate the line-integral along a circle of latitude and use it to calculate the earth-air current remembering that the value of the integral is equal to 4π times the current enclosed. The calculated current intensity is found to be more than 10^4 times as great as the earth-air current measured by the usual apparatus for studying atmospheric electricity. The origin of this discrepancy is not yet known. The other part of the external field which possesses a potential has its origin in overhead current systems. It is the variation of these currents which produce the transient terrestrial magnetic variations and will form the main theme of this chapter.

The origin of the main geomagnetic field of the earth is still ill-understood. A very promising hypothesis had been put forward by Blackett [2] some time ago based on some earlier suggestions by Schuster [2*a*], Sutherland [3] and others. According to this hypothesis a fundamental property of matter, perceptible only in bodies of great size and mass, is that when rotating it develops a magnetic moment, the ratio between the angular momentum and the magnetic moment being of the order $cG^{-\frac{1}{2}}$ (c—velocity of light, G—gravitational constant). This ratio agrees well for the case of the earth. The hypothesis received support from the fact that the magnitude, as well as the direction of the general magnetic field of the sun, as per estimate of Hale made long ago, is related to the angular momentum as demanded by the hypothesis. It was also in conformity with Babcock's [4] discovery of large magnetic fields in rotating stars. However, according to contemporary measurements by Thiessen [5] it is found that the general magnetic field at the surface of the sun is of the order 1 gauss only, instead of 50 gauss as estimated by Hale and that too in the opposite sense. Further, a strong magnetic star has been discovered which changes the polarity of its magnetization every nine days [6], and it is most unlikely that such changes are concurrent with changes in the sense of rotation. These facts cast doubt on the otherwise attractive theory of Blackett.

Another theory, entirely on classical lines, has been developed by Bullard [7] and by Elsasser [8]. The earth's magnetic field is ascribed to electric currents flowing in the metallic fluid core, and maintained by induction effects caused by the motion of the fluid in the magnetic field. The process resembles that occurring in a self-exciting dynamo. The source of the fluid motion may be ascribed to thermal convection due to localized production of heat by radioactivity. Bullard [9] has also outlined a theory of secular variation. It is supposed that there are large scale eddies near the surface of the earth's liquid core and that the electromotive forces induced by these eddies in the presence of the permanent magnetic field of the earth drive electric currents in this part of the core. As the eddies (of very long period) appear or disappear the electric currents and the magnetic field produced by them also vary, and this is observed on the surface of the earth as secular variation.

(c) The magnetic elements

In discussing the variations of the magnetic elements, it is convenient to represent the elements by symbols. A list of such symbols commonly employed is given below:

- F—total magnetic intensity
- H—horizontal component of the magnetic intensity
- Z—vertical component of the magnetic intensity—reckoned positive when downwards
- V—same as Z regardless of sign
- X—northern component of H
- Y—eastern component of H
- W—western component of H ($= -Y$)
- I—magnetic dip or inclination; the angle by which a freely pivoted magnetic needle dips below the horizontal—reckoned positive when the north-seeking pole of the needle points downwards and negative when it points upwards

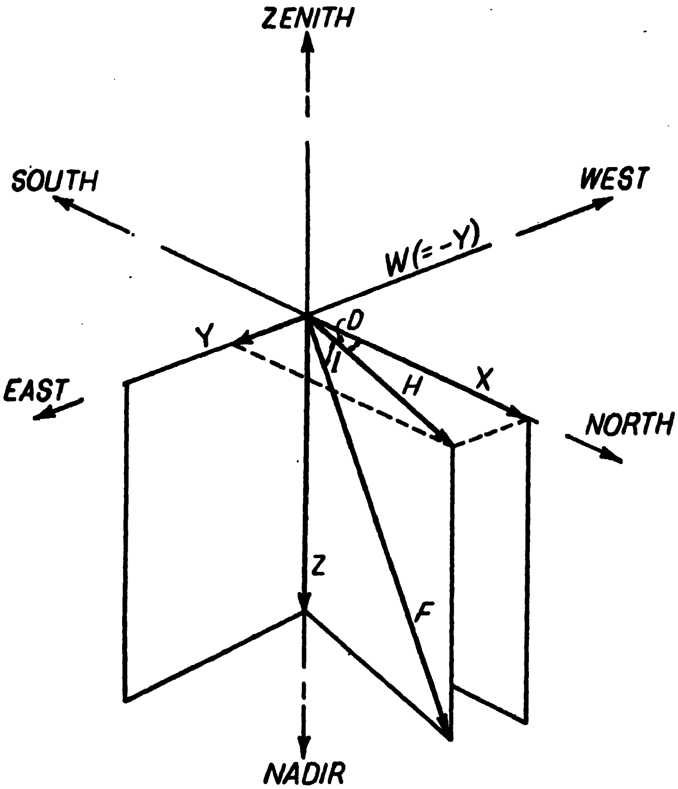


FIG. 2. Illustrating the different magnetic elements:

- F—Total magnetic intensity
- H—Horizontal intensity
- X—Northern component
- Y—Eastern component
- Z—Vertical component
- D—Declination
- I—Dip
- W—Western component

D—declination; the angle between the vertical plane through the axis of the magnetic needle and the geographical north-south meridional plane—reckoned positive if the former is to the east of the latter. It is called 'variation' by the mariners.

The above elements are schematically represented in Fig. 2.

The relations between the various elements are given by

$$\left.\begin{aligned} F &= H \sec I \\ X &= H \cos D \\ Y &= H \sin D \\ Z &= H \tan I \end{aligned}\right\} \quad \cdots \quad (2)$$

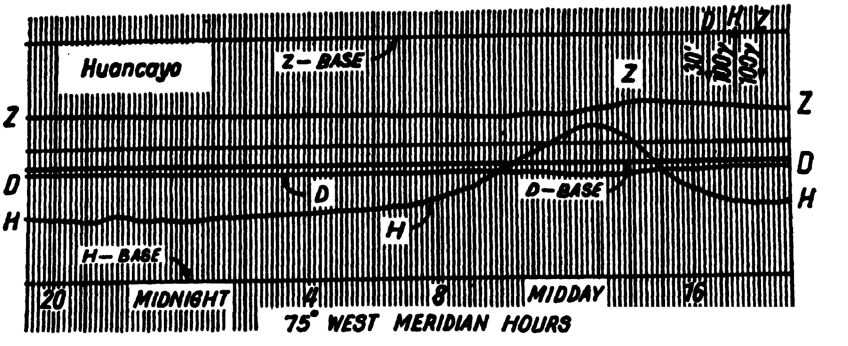
Of the various elements, only three are required to specify the field at a point both in magnitude and in direction. It is therefore sufficient for the magnetic observatories to record any three of these quantities.

2. TRANSIENT VARIATIONS OF THE MAGNETIC ELEMENTS

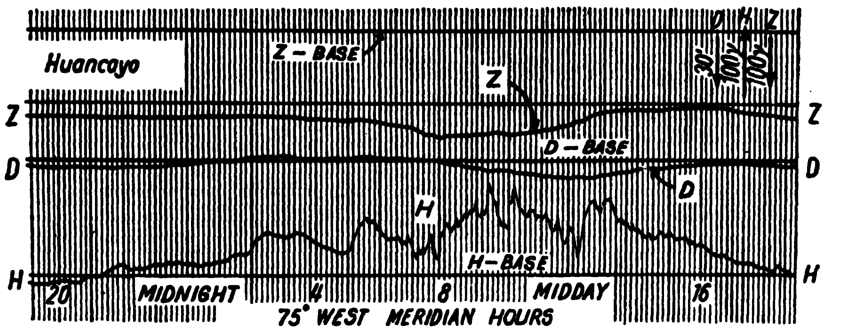
(a) Quiet and disturbed days—Magnetic character figure—Three-hour range indices

(i) *Magnetic character figures.*—The three magnetic elements of which records are usually kept are H, the horizontal intensity, Z, the vertical intensity and D, the declination. Records of any magnetic observatory show that on some days these three elements undergo smooth and regular variations, while on others they are more or less disturbed. For a long time it has been customary according to an international scheme to characterize each day at any observatory by a so-called *magnetic character figure* (C) for the day. Days of quiet magnetic condition are given the figure 0; days of strong disturbance are given the figure 2, while a day of moderate disturbance is characterized by the figure 1. In Fig. 3 magnetograms of H, Z and D recorded at Huancayo on internationally quiet and disturbed days are reproduced.

This method of classification by simple inspection of records, crude as it is, has proved to be of great value in many magnetic investigations. The daily figures 0, 1, 2 from all the observatories are collected at a central office (De Bilt, Holland) where the average for each day is taken. The average figure, to one place of decimal, which lies between 0 and 2, is called the *international magnetic character figure* C_i for the day. From the international character figures for each day of the month the central office selects, for each calendar month, five days which appear to be the quietest and five which appear to be the most disturbed. These days are called the *international quiet* and *international disturbed* days respectively. The use of a daily planetary character figure C_p has also been proposed by Bartels [9a]. This figure is to be regarded as a daily index for the intensity of the solar corpuscular radiation for the entire earth. The scale for C_p will be the same as that for C_i, ranging from 0·0 to 2·0, with the option to split 2·0 into 2·0 to 2·5.



(A)



(B)

FIG. 3. (A) Magnetograms showing the variations of H, Z and D at Huancayo on August 8-9, 1929, a quiet day of international character figure 0·0. (B) Magnetograms showing the variations of H, Z and D at Huancayo on December 3-4, 1929, a disturbed day of international character figure 1·6.

(ii) *Magnetic activity—the u- and u_1- measures.*—The above method of classifying days by assigning character figures, though very useful, is not always satisfactory as it depends on the subjective judgment of the observer in charge at the particular observatory. An objective measure of the intensity of magnetic disturbance or *magnetic activity*, independent of any judgment of the observer is therefore necessary. In recent years such an objective measure of magnetic activity has come into use. It will be seen later that a singular feature of a magnetic disturbance is that the average value of the horizontal intensity is depressed immediately after the incidence of the disturbance. This phenomenon, called post-perturbation, is utilized for measuring the magnetic activity.

The nature of the post-perturbation will be understood from Fig. 4. Each of the curves shows how the value of H is depressed during a magnetically disturbed period and how it gradually recovers its normal value thereafter. The similarity of the three curves is remarkable and shows that the change in the average value of H from one day to the next can be used as a universal measure (u) of the magnetic activity for the period

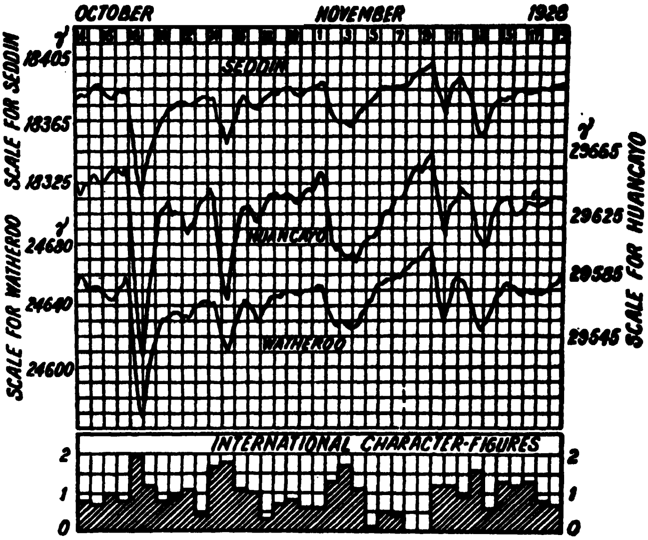


FIG. 4. Depression of H following the incidence of magnetic disturbance and gradual recovery to the normal value. The curves show the variation of 24 hourly mean values. The vertical lines represent Greenwich midnight. (After Chapman.)

under consideration [10]. The computation of u is done as follows. To each day is assigned the difference of the mean value of H (or of X) from that of the preceding day. The average of these differences over a month may be called the mean interdiurnal variability U for the month. Now, as we shall see later, a world-wide magnetic disturbance may be considered as due to circular current systems in the outer space encircling the earth parallel to the magnetic latitudes. The magnetic field of such a current system is directed along the magnetic axis and only its component in the direction of H (or of X) is effective in producing the observed changes in these elements during a magnetic storm. Hence we introduce the universal magnetic activity u given by

$$u = U/\cos\beta,$$

where β is the angle in space between the directions of the magnetic axis and H (or X).

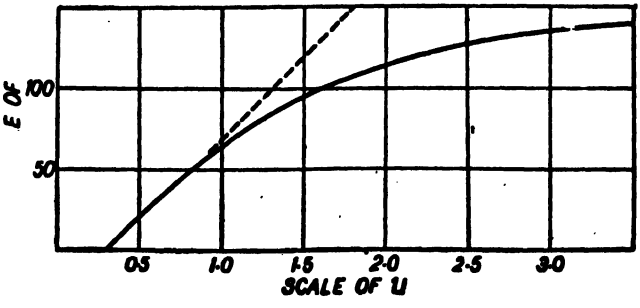


FIG. 5. Relation between u and u_1 measures. (After Bartels.)

Another measure of the magnetic activity is the u_1 measure which is a function of the monthly mean of u. Since the value of the daily character figure C cannot exceed 2, it follows that the great magnetic storms get less weight in the monthly means of the character figure C than in the monthly means of u which have got no upper limit. It is, however, sometimes desirable to have a measure of magnetic activity which is just as well defined as u but at the same time throttles down somewhat the influence of exceptionally great disturbances. Such a measure is u_1 introduced by Bartels [10]. Fig. 5 shows, after Bartels, how the u_1 measure is related to the u measure.

(iii) *The K-indices.*—The magnetic variations as recorded on the earth are partly 'regular' and partly 'irregular'. The 'irregular' variations have a special importance in that they are supposed to indicate the influence of solar corpuscular radiation on the Earth. In order to provide a homogeneous running record of these variations (and hence of the solar corpuscular radiation) a three-hour range index K has been introduced [10a]. The K variations include the regular fluctuations in the magnetic elements during disturbance (D), the mean effect of the disturbance-field (D_m) as manifested in changes throughout the three-hour interval and the disturbance daily variation (S_D) including that part of D_m called the non-cyclic variation on disturbed days. The principle and practice of scaling K are as follows: [For the definitions of the symbols D, D_m and S_D see Sec. 2(b)].

Each observatory assigns to each of three-hour intervals beginning at the Greenwich mean time 0^h, for the value of K, one of the integers 0 to 9. For each observatory a permanent scale is adopted which gives the limits which the ranges R, measured in units of force γ, define the index K. For each of the three magnetic elements (D, H, Z) or for the rectangular field-components X (northward), Y (eastward) and Z (downward), the range R is defined as the difference between the highest and the lowest departure (during the three-hour interval) from the regular daily variation expected for that element on a magnetically quiet day, according to the season, the sunspot cycle and, in some cases, the phase of the moon. The value of R for the most disturbed element is taken as the basis of the index K.

The K-scales are determined in such a way that each observatory has, within a year (for example), approximately the same number of intervals with $K=0$ or with $K=1$, etc. The K-scale based on the '*Potsdamer erdmagnetische Kennziffer*' which has proved satisfactory may, in this connection, be described. This is

$K =$	0	1	2	3	4	5	6	7	8	9
$R =$	.. 5	.. 10	.. 20	.. 40	.. 70	.. 120	.. 200	.. 330	.. 500 γ	

This means that all ranges R smaller than 5γ give $K=0$; all ranges between 5γ to 10γ give $K=2$ and so on. It will be noted that the scale proceeds by multiples of 2 up to $R=40\gamma$ and then slows down; otherwise the higher indices of K will never be reached.

It has recently been recommended [9a] that for non-polar stations (with limits for $K = 9$ below 500 γ), it would suffice to combine $K = 0$, 1 and 2, into one step $K = Q$, so that the K-scale for these stations will have the steps $K = Q$, 3, 4, 5, 6, 7, 8, 9.

The K index is only a 'regional' index. It gives a measure of the intensity of geomagnetic disturbance of the place concerned and is not free from local features. There is, however, a need for a measure of the disturbance for the earth as a whole. Such a measure is provided by the three-hour index K_p called the geomagnetic 'planetary' index. It is based on 'standardized' indices K_s which are freed from local features as far as possible by an appropriate standardizing process. So far conversion-tables assigning a K_s-index to every K-index have been derived for 11 observatories. K_p is the average of the eleven K_s-indices.

(b) The different types of transient variations

(i) *Observatory data and their analysis: the S-variations.*—The *hourly* values of the magnetic elements measured at the observatories are the starting point for computing the various types of transient variations. Usually these hourly values are the mean values of the magnetic elements for one hour intervals which extend either from one exact hour to the next or are centred at each exact hour.

Daily mean value.—The mean of the 24 hourly values from midnight (0 h.) to the following midnight (24 h.) is called the *daily mean value.*

Monthly (or annual) mean value.—The daily mean values, if averaged over a month (or year) is called the *monthly (or annual) mean value.* In the same way *group mean values* for a group of selected days may be obtained from the daily means.

Mean hourly value.—The mean of the hourly values for *each individual hour* for *all* the days of a month (or a year) is called the *mean hourly value* (of the individual hour) for the month (or the year). Similarly, mean hourly values may be obtained from a selected group of days (instead of from *all* the days) of a month. Such selected groups are usually the international five quietest and five most disturbed days of the month.

The sequence of the 24 mean hourly values thus derived indicates the mean daily variation of the element for the month (or the year or the selected group of days).

Mean daily inequality (S).—The monthly (or yearly) mean value subtracted from the mean hourly values of the month (or the year) gives a sequence of *hourly departures* which is called the *mean daily inequality* for the month (or year) and is denoted by S.

Similarly the mean daily inequality may also be obtained for a particular group of days.

Daily variation curve.—The 24 hourly departures plotted against time is called the *daily variation curve* of the element for the month (or the selected group of days).

S_q.—The mean daily inequality derived from the five quietest days of the month is called S_q.

S_d.—The mean daily inequality obtained from the five most disturbed days of the month is called S_d.

The method of deriving the various quantities and types of variations defined above may best be illustrated by taking an idealized chart of hourly values as in Table I.

TABLE I

HOURLY VALUES.

Hour→ / Day ↓	1 h.	2 h.	3 h.			22 h.	23 h.	24 h.
1	1_1	2_1	3_1	....	....	22_1	23_1	24_1
2	1_2	2_2	3_2	....	....	22_2	23_2	24_2
3	1_3	2_3	3_3	....	....	22_3	23_3	24_3
..	..	..	..	....	....	..	..	..
30	1_{30}	2_{30}	3_{30}	....	....	22_{30}	23_{30}	24_{30}

From the hourly values tabulated above, the *mean hourly values* for the month and the *daily mean values* are obtained as shown in Table II.

TABLE II

Hour.	Mean hourly values (for the month).	Day.	Daily mean values.
1 h.	$\frac{1_1+1_2+1_3+..+1_{30}}{30}=h_1$	1	$\frac{1_1+2_1+3_1+..+24_1}{24}=d_1$
2 h.	$\frac{2_1+2_2+2_3+..+2_{30}}{30}=h_2$	2	$\frac{1_2+2_2+3_2+..+24_2}{24}=d_2$
..	....	..	....
24 h.	$\frac{24_1+24_2+24_3+..+24_{30}}{30}=h_{24}$	30	$\frac{1_{30}+2_{30}+3_{30}+..+24_{30}}{24}=d_{30}$

$$\textit{Monthly mean} = \frac{d_1+d_2+d_3+\ldots\ldots\ldots\ldots+d_{30}}{30} = d_m$$

Mean daily inequality for the month, S = Sequence of 24 hourly departures $\begin{cases} h_1-d_m; \\ h_2-d_m; \\ h_3-d_m; \\ \cdots\cdots \\ h_{24}-d_m. \end{cases}$

Daily variation curve for the month is the plot of (h_1-d_m), (h_2-d_m), (h_3-d_m), ...$(h_{24}-d_m)$ against time.

The above quantities may also be computed for all the days of a year or for a group of selected days instead of for a month. Thus we may select from the monthly chart five quietest days or the five most disturbed days of the month. The hourly values of these groups of days may then be utilized in the above manner to find out the *group mean value*, the *daily mean inequality* or the *daily variation curve* for either of these two groups of days.

(ii) *Disturbance variations.*—The disturbance variations are: S_D, (*disturbance daily variation*), D_{st}, (*storm-time variation*) and D_i, (*irregular variations*).

Disturbance daily variation (S_D).—There is a systematic difference (particularly noticeable in higher latitudes) between S, the mean daily variation obtained from all days of a month or S_d obtained from the five most disturbed days of a month and S_q the mean daily variation obtained from the five quietest days. This difference $S-S_q$ or S_d-S_q is caused by magnetic disturbance D and is due to that part of the disturbance, which is related to the time of the solar day. The sequence $S-S_q$ (or S_d-S_q) is called the *disturbance daily variation* and is denoted by S_D.

D_m.—Besides S_D as defined above, the magnetic disturbance has two other components. One of them modifies the daily mean value of the elements, e.g., decreases H. This part can therefore be obtained by subtracting the mean values of the elements on quiet days from the mean values on disturbed days or all days. The difference is obviously greater in the former case than in the latter. This difference gives the average change in the mean value of the elements due to the part of D which is denoted by D_m. It will be presently seen that D_m is, on the average, the daily mean of D_{st}, the storm-time variation.

Storm-time variation (D_{st}).—The variations accompanying a magnetic disturbance are not wholly irregular but have certain regular features. This regular part of the disturbance is the storm-time variation D_{st} (Fig. 9). The nature of D_{st} variation curve may be revealed in the following manner.

At a particular station a number of magnetic storms of roughly similar intensity is chosen such that the times of commencement of the storms are fairly evenly distributed over the 24 hours of Greenwich time. A particular magnetic element is now chosen and its successive hourly values are written out in rows (one row for each storm), each row commencing with the value of the element for the hour preceding the commencement of the storm. The hourly values in the vertical columns are added and means taken. In the 24 average hourly values thus obtained, all periodic changes of the magnetic element which depend on the local time, e.g., S_q or S_D will evidently cancel out owing to the fact that the hours of commencement are nearly uniformly distributed over the solar day. The variation with time of the magnetic element, indicated by the final row of means, represents the storm-time magnetic disturbance D_{st}.

It is now evident that the average difference between the daily mean value of a magnetic element on disturbed (or all) days and the daily mean value on quiet days, i.e., D_m is in fact equal or proportional to the daily mean of D_{st}.

D_i.—The remainder of D is wholly irregular being neither related to storm-time nor to solar time. This irregular part of D is denoted by D_i. Obviously, it can be obtained by subtracting S_D and D_{st} from D.

(iii) *Lunar diurnal variation L*.—Besides the variations mentioned above there is a regular semi-diurnal variation connected with the lunar day. The magnitude of this variation is very small—about one-tenth of the solar variations—and cannot be detected by mere visual inspection of the records of the magnetic elements of a single day as is the case with the solar daily variation. In order to make reliable determination of L material for many years is necessary. The computation can be done in the following way.

The monthly mean value of the magnetic element is subtracted from the mean hourly values for the month, giving a set of 24 hourly departures. From the value of the element at each hour during the month, the departure for that hour is subtracted, leaving a set of hourly values free from solar daily variation. These hourly values will vary because they contain L and also because of magnetic disturbance. A large number of sets of these 24 successive hourly values are now arranged in rows, each commencing with the hour nearest to that of the lunar transit. If now the means of the hourly values in each column are taken then the variations due to the magnetic disturbance will cancel out as they are not related to the lunar time. The final row will thus give only that part of the variation of the earth's field which is related to the lunar time, i.e., the lunar daily variation L.

General statistical methods for the determination of S and L have been described in full by Chapman and Miller [11].

The classification of the transient magnetic variations described above will be clear from the chart below:

TRANSIENT TERRESTRIAL MAGNETIC VARIATIONS

- Quiet Day Variations
  - S_q Solar Diurnal Variation
  - L Lunar Diurnal Variation
- Disturbance Variations
  - S_D Disturbance Daily Variation
  - D_{st} Storm-time Variation
  - D_i Irregular Variation

(iv) *Dependance on latitude of the average daily variation curves.*—The average shape, phase and amplitude of the variation curves discussed above

depend on the latitude of the place of observation. (The amplitude of the average daily variation curves also depends on the season of the year and on the epoch of the solar cycle. These will be referred to later.)

In Figs. 6, 7 and 8 the S_q- and S_D-curves for the magnetic elements H and V and W-declination respectively are shown for a number of latitudes [12]. The curves 1 to 5 refer to the following observatories or groups of observatories:

(1) Sitka, Mag. Lat. 60°.
(2) Pavlovsk, Mag. Lat. 56°.
(3) Pola, Potsdam, Greenwich, mean Mag. Lat. 51°.
(4) Zikawei, San Fernando, Cheltenham, Baldwin, mean Mag. Lat. 40°.
(5) Batavia, Porto Rico, Honolulu, mean Mag. Lat. 22°.

In each of the figures, column (a) depicts S_q while (b) and (c) depict S_D derived respectively from disturbed days, i.e., from S_d, and from all days, i.e., from S. It may be noted that on disturbed days the quiet day variations shown in (a) are superposed on those shown in (b).

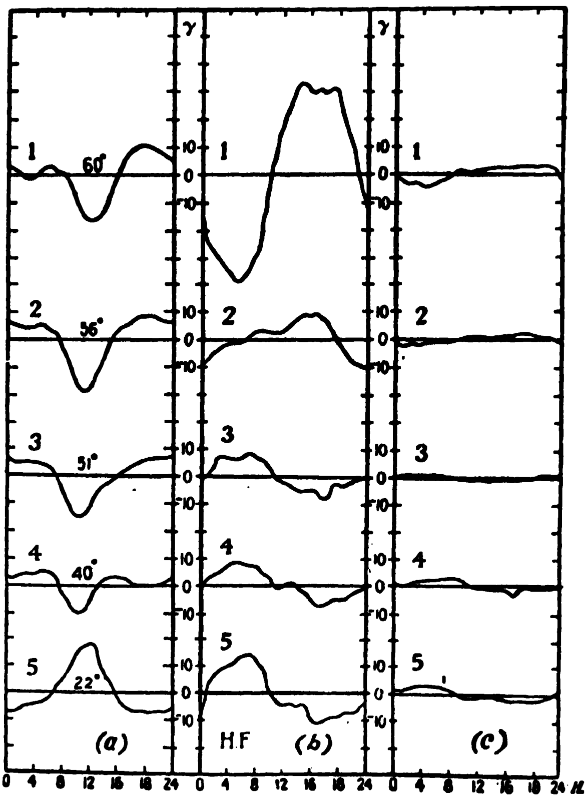


FIG. 6. Average daily variation curves of H at different magnetic latitudes (a) S_q; (b) $S_D = S_d - S_q$; (c) $S_D' = S - S_q$. (After Chapman.)

Fig. 9 shows the storm-time changes at three groups of observatories indicated in the figure [13]. The difference in the hours plotted in the abscissa in the two sets of diagrams, Figs. 6–8 and Fig. 9 should be noted. In Figs. 6–8 the abscissa is local time and the origin is 0 hour local time; in Fig. 9 the abscissa is storm-time and the origin is the time of commencement of the storm.

Fig. 10 shows the lunar diurnal variation of the magnetic west component at Batavia, derived as explained above [14]. It is seen that for all

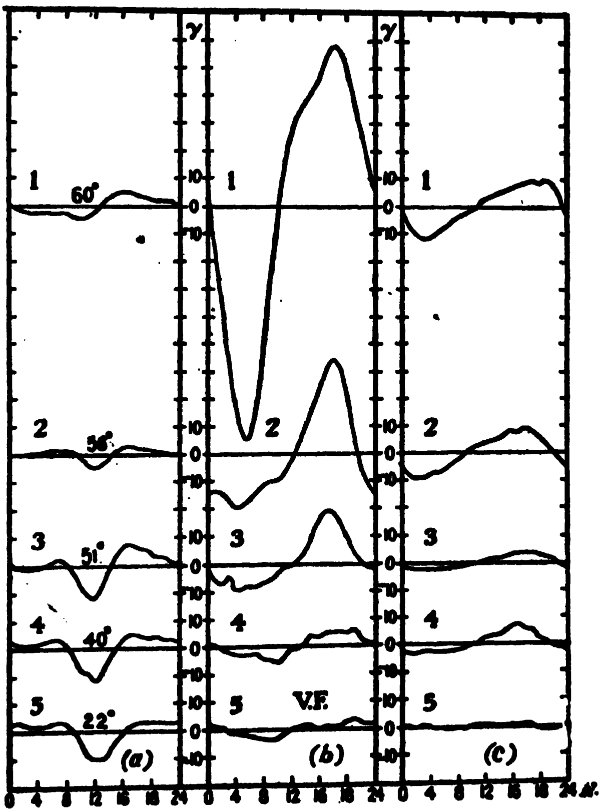


FIG. 7. Average daily variation curves of V at different magnetic latitudes: (a) S_q; (b) $S_D = S_d - S_q$; (c) $S_D = S - S_q$. (After Chapman.)

phases of the moon the variation is much more pronounced during the daylight hours (thicker portions of the curves) than at night. The mean variation averaged over all the phases is a simple sine curve of period 12 hours. The variation for each individual phase is not so simple. (The average L-variation also depends on the latitude of the place of observation, though, curves depicting same are not given. The L-variations will be further discussed in Sec. 4.)

The distinctive features of the different types of transient variations may thus be summarized: The magnitude of the variations is always

small being only a few thousandths or a few ten-thousandths of the permanent magnetic field. The L-variations are about one-tenth of the S-variations. The periodic variations are related either to the solar time or to the lunar time, i.e., for an observer situated on the sun (or on the moon) the distribution of the solar (or the lunar) variations over the surface of the

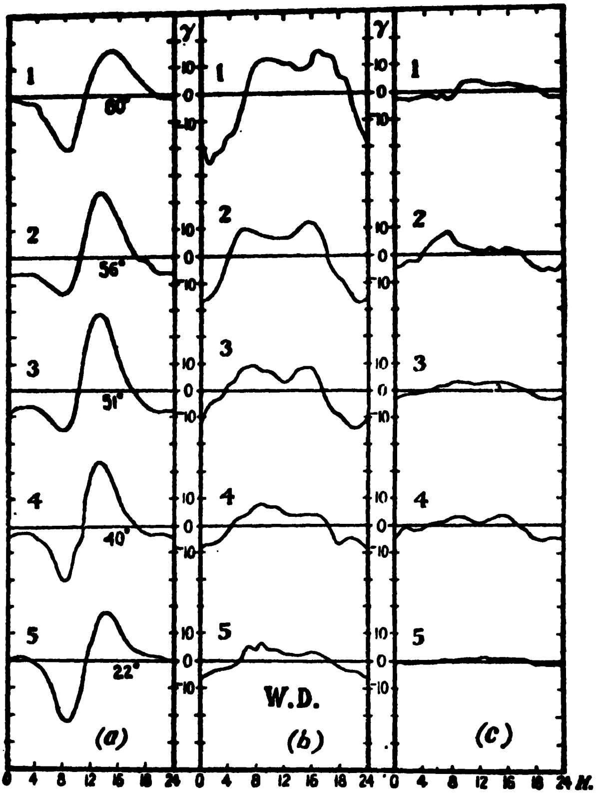


FIG. 8. Average daily variation curves of W. Dec. (in force units) at different magnetic latitudes: (a) S_q; (b) $S_D = S_d - S_q$; (c) $S_D = S - S_q$. (After Chapman.)

earth appears more or less constant with time. In contrast to the S and L variations the storm-time variation is related to the so-called storm-time. The storm-time field is symmetrical with respect to the earth's magnetic axis.

3. THE QUIET DAY SOLAR VARIATION S_q

(a) The main features of S_q variation

Fig. 6 column (a) shows some of the important features of the quiet day solar variation of the horizontal force. We notice, for instance, that the amplitude of the variation is generally greater during daylight hours than during the night and that the variation changes phase at a latitude somewhat below 40°. The amplitude of S_q has also an annual variation

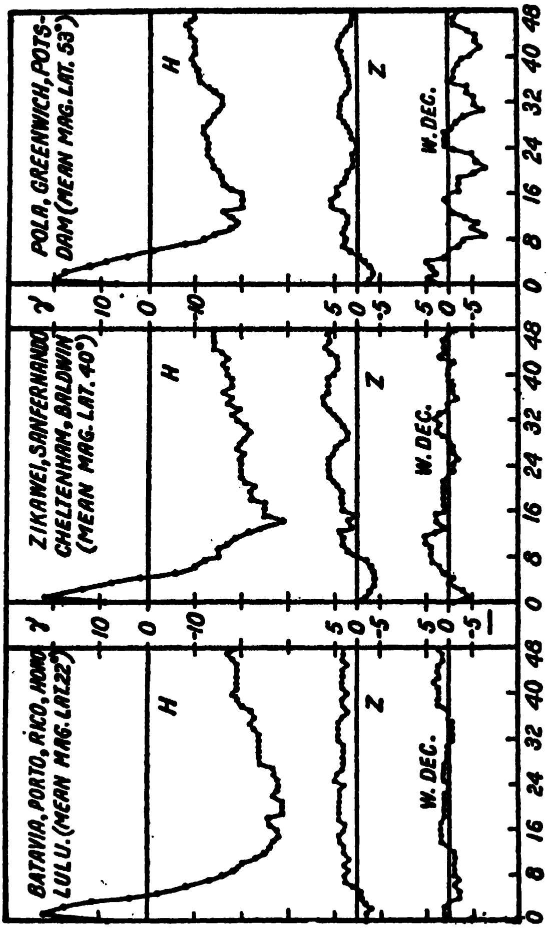


FIG. 9. Average variation in the magnetic elements during storm-time at different latitudes. (After Chapman.)

Erratum: For the mean magnetic latitude of the group of observatories, Pola, Greenwich, etc., read 51° instead of 53°.

(not shown in the figure) which is particularly noticeable in moderate and high latitudes; the amplitude is greater in summer than in winter. Further, the intensity of S_q is found to vary appreciably from day to day even during periods which are magnetically quiet. This day-to-day variability has been discussed in considerable detail by Hasegawa [15].

(b) Variation of S_q with the sunspot cycle

The amplitude of S_q also varies with the sunspot cycle. It is notably greater in years of sunspot maximum than in years of sunspot minimum. The close correlation between the range of S_q and the annual mean sunspot numbers is shown in Fig. 11 where curve R refers to the range of the mean annual S_q variation of H at Bombay, India, and curve S refers to annual mean sunspot numbers. It will be noticed that the variation of R, though not strictly parallel to the variation of the sunspot number S, closely followed the same throughout the eleven-year cycle 1894–1905.

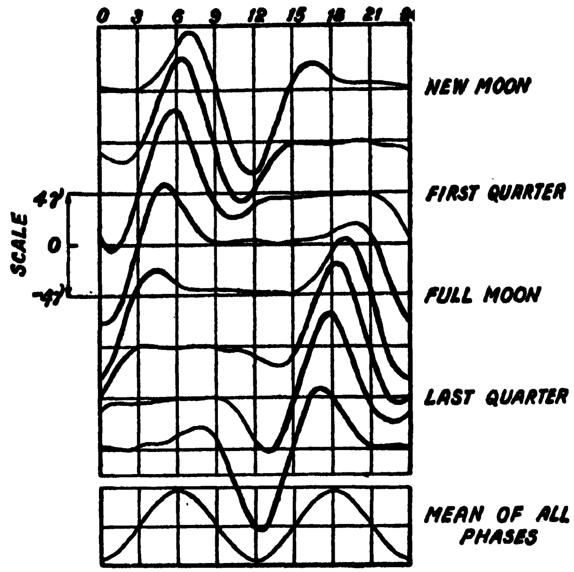


FIG. 10. Lunar diurnal variation of magnetic west component (in force units) at Batavia for different phases of the moon and the mean variation averaged over all the phases. (After Chapman.)

We may thus summarize the facts regarding S_q by saying that it is larger over the sunlit than over the dark hemisphere; that it is also so over the summer than over the winter hemisphere; and that the intensity of the S_q-field varies with the sunspot cycle and is from 50 to 100 per cent greater at sunspot maximum than at sunspot minimum.

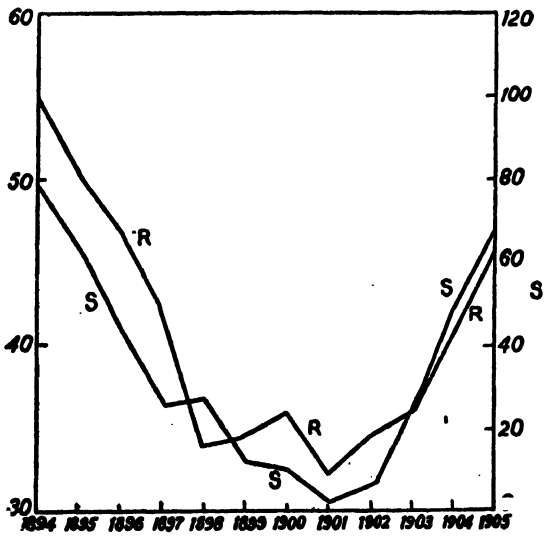


FIG. 11. Illustrating the parallelism between the range (R) of annual mean S_q of the horizontal force (at Bombay) and annual mean sunspot numbers (S).

(c) The atmospheric current system for the S_q-field

Analysis of the S_q variations by a method first indicated by Gauss [16] and later developed by Schuster [17] shows that the S_q-field possesses a potential, i.e., the field is produced by magnetism or electric currents existing above or within the earth. It further shows that the origin of the S_q-field lies neither wholly below nor wholly above the earth's surface. The major portion—about two-thirds to three-quarters—of the total field near the surface of the earth has its origin above it and is called the external part S_q^e. A minor portion has its origin below the surface and is called the internal part S_q^i.

It is found that the external and the internal field systems which together produce the observed S_q variations on the surface of the earth are not unrelated to one another. Harmonic analysis of the S_q^e and S_q^i fields shows that the amplitude ratios and the phase differences of their different periodic components are similar. This suggests that there is a causal relation between the S_q^e and S_q^i fields and, since the former is much greater than the latter, the external part S_q^e should be the cause and the internal part S_q^i should be its effect. There is also an obvious way of associating the internal field with the external field. Any variation in the agency causing S_q^e can, by magnetic induction, produce secondary current systems inside the earth. In fact, if we assume that the interior of the earth has a uniform conductivity of $3{\cdot}6\times10^{-13}$ e.m.u. throughout a con centric core of radius 4 per cent less than the earth's radius and that the outer 4 per cent of layer of thickness 260 km. is non-conducting, then the observed amplitude ratios and the phase differences between the various components of the S_q^e and S_q^i fields can be explained as due to the induction of the latter by the former [18, 19].

From what has been said above it is evident that the origin of the external part S_q^e is the ultimate origin of the S_q-field as observed at the earth's surface. Since the atmosphere is practically non-magnetic and non-conducting up to about 50 km., we can conclude that the origin of the S_q^e must be above this height. The presence of conducting ionospheric layers above this height provides a possible medium for current systems which, together with the current systems induced by them inside the earth, may produce the observed S_q variation. The nature of the external current system which might produce the observed S_q variation can be computed, independent of any theory regarding the exact manner by which the current system is produced. This has been done by Bartels [20] by applying the laws of electromagnetism and utilizing the results of spherical harmonic analysis of the S_q variations made by Chapman [21]. The results of the calculations are shown in Figs. 12 and 13. Fig. 12 shows the current distribution on a Mercator projection during the equinoctial period and Fig. 13 that in June. The meridians of longitude are not fixed with respect to the earth but are so relative to the sun. The current system will therefore appear stationary to an observer on the sun. It is clear that as the earth revolves,

a point on a given latitude is brought successively under different parts of the current sheet in this latitude and experiences a corresponding diurnal variation of magnetic force.

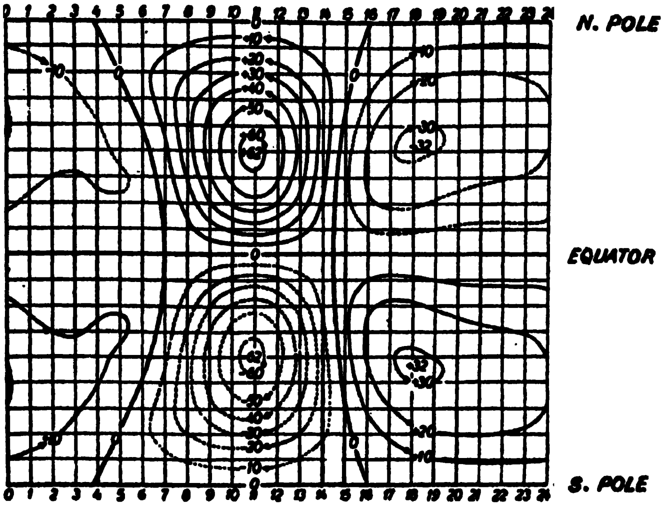


FIG. 12. Upper atmospheric current system at a height of about 100 km. which may produce the observed S_q variation of magnetic elements during the equinoxes. (After Chapman and Bartels.)

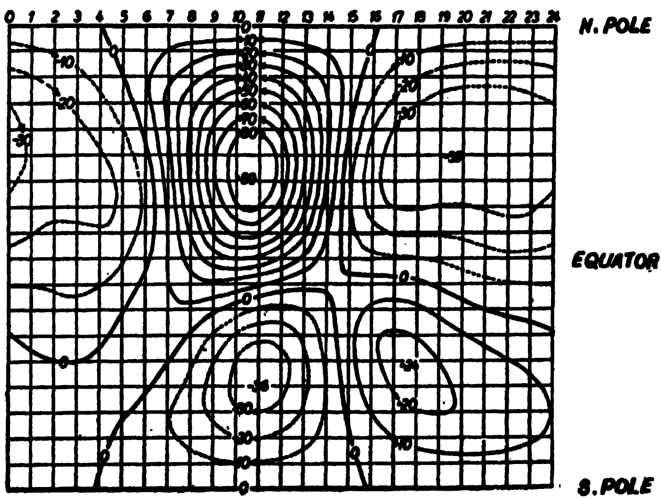


FIG. 13. Upper atmospheric current system at a height of about 100 km. which may produce the observed S_q variation of the magnetic elements in June. (After Chapman and Bartels.)

The arrows indicate the directions of the current flow and the lines are drawn such that between each pair of adjacent lines a current of 10,000 amperes flows. There are four current systems, two to the north and

two to the south of the equator, each flowing round a centre of focus. For the equinoxes (Fig. 12) the current system is symmetrical with respect to the equator. For the sunlit hemisphere the foci do not lie on the noon meridian but are removed about 2 hours from it. It may be noted that the equatorial current intensity per cm.′ across the meridian has an eastward maximum of 2×10^{-5} e.m.u. at 11 h. and a westward maximum of 10^{-5} e.m.u. at 17 h. The total current flow is much greater in the sunlit portion than in the dark portion of the earth, so that the S_q-field is greater at day than at night. The current system shown in Fig. 13 for June is not so symmetrical as that of Fig. 12 for the equinoxes. This is because at the equinoxes the northern and southern hemispheres are equally lighted by the sun, whereas in June, i.e. near the summer solstice, the northern hemisphere is lighted up to a greater extent than the southern hemisphere. The total current flow in the day-circuit in the northern hemisphere is 89,000 amperes in June but is reduced to 62,000 amperes at the equinoxes.

Results of measurements of total magnetic field intensity at high altitude appear to show that the S_q current system flows in the E-region [see Chap. XII, Sec. 4(g)]. The same conclusion is also arrived at by Terada [22] who has constructed the S_q current system by a simple method. Martyn, however, considers that the main current is in the D-region [22a].

4. THE LUNAR DAILY VARIATION L

We have already noted that the variations of the magnetic elements contain a part which follows the lunar day and is called the lunar daily magnetic variation L. The amplitude of L is much smaller than that of S_q and the procedure by which it can be isolated from other variations of much larger magnitude has been indicated. We now describe briefly the salient features of the L-variation.

(*a*) The L-variation

If the L-variation is derived from all days in a particular month for a number of years or for all days in a year, it is found to be very simple. The curve representing this monthly mean L is a double sine-wave as mentioned before. One can also derive L from a large number of days all at a particular phase of the moon such as the new moon, full moon, etc. When so determined the curve is not so simple as the curve for all days of the month or the year. It is found that the portion of the curve which corresponds to daylight hours has greater range than that of the remaining portion which corresponds to the hours of darkness. Fig. 10 represents a series of such curves drawn for successive phases of the moon [14]. Since during a whole lunar month or year, the daylight hours are equally distributed over all parts of the lunar day, the light to darkness inequality is cancelled out in the mean L obtained by averaging the variations for all the phases in a

lunar month and a double sine-wave is obtained. This is shown by the curve at the bottom of Fig. 10.

Just as the quiet day solar variations S_q increases in amplitude from winter to summer, so also the amplitude of L shows a seasonal variation. However, the summer to winter ratio of L is much greater than that of S_q. An inspection of Figs. 15, 16 to be described presently shows that the total current in the main circuit of the overhead current system for L at the equinoxes is about 5,300 ampères whereas in the summer it is about 11,000 amperes, the ratio being about 1 : 2·6.

At Huancayo on the magnetic equator (South America) the lunar diurnal variation in H is of large magnitude, not only in absolute units but even relative to S, which itself is known to be exceptionally large at this station. This may be compared with other stations where, as already mentioned, L is so small that it can be extracted only from many hourly values.

Bartels and Johnston have taken advantage of this special characteristic of the magnetic variations at Huancayo for determining the geomagnetic tidal effects in H [23]. They selected from the magnetogram records of the period 1923–1939 days with exceptionally large values of L (the 'big-L-days') and determined therefrom the lunar semi-monthly waves in the range of H for every hour of the solar day (see Fig. 14). Striking day to day changes in L have also been detected. But, L and S fluctuate rather independently of each other and the relative fluctuations of L appear to be larger than those of S. As expected, the lunar semi-monthly waves in the range of H change with the season and with the solar cycle. There is, however, some difference in the changes as compared to those of S.

(b) Variation of L with sunspot and magnetic activities

The observed increase in the range of L in daylight hours and also in summer naturally suggests that, like S_q, L will also vary with the sunspot cycle. In order to verify this L is determined for groups of years having roughly the same mean sunspot number. For comparison, S_q is also determined for the same groups of years. The range of L and also that of S_q are then plotted against the mean sunspot number. It is seen that while the range of S_q increases about two-fold from years of least to those of greatest sunspot number that of L increases only slightly.

In the work of Bartels and Johnston [23] at Huancayo mentioned above it was found that in the months November to March (when L is larger than in the rest of the year) both S and L increase proportionately to each other from sunspot minimum to sunspot maximum. But, around June (when L is small) it does not increase (as the S does) in the same manner from the minimum to the maximum.

We have already seen that during periods of magnetic disturbances, the solar daily variation consists of two parts S_q and S_D. Of these two, S_D varies in intensity (but not in form) in accordance with the degree of

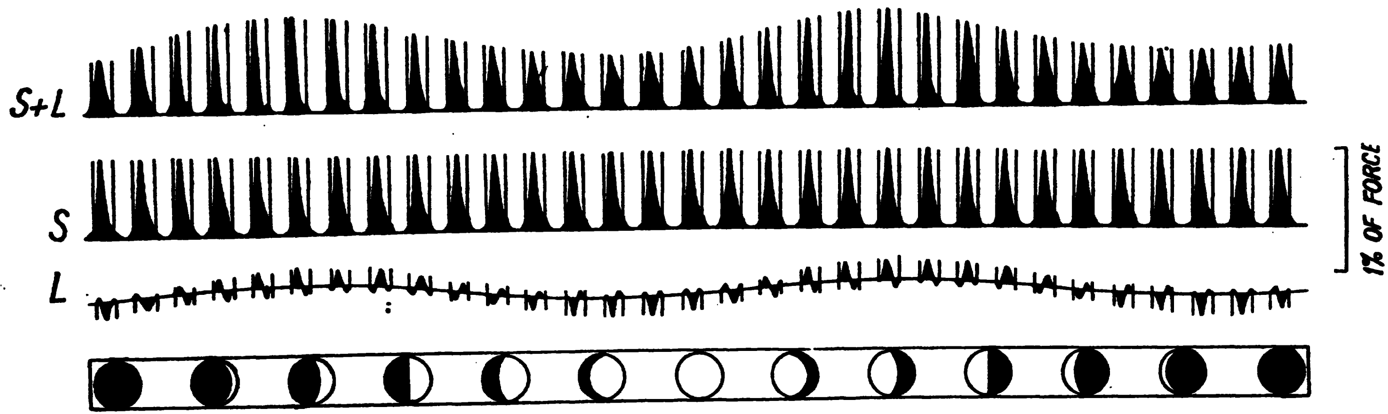


FIG. 14. Illustrating diurnal changes in one lunar month in the range of H caused by solar and lunar tidal motions at Huancayo (the solid curves). Advantage has been taken of the fact that the tidal effects at this station are exceptionally large. In the uppermost set of curves the solar and lunar tidal effects are superposed. In the second set the solar diurnal variations only are depicted. In the third set the lunar diurnal variations only are depicted. The phases of the moon are shown at the bottom. The computations are for a period of sunspot maximum. Similar computations made for a period of sunspot minimum show that the amplitudes of the variations are much smaller. (After Bartels and Johnston.)

the disturbance. In a similar manner L also increases in intensity as the disturbance increases but the increase is dependent on the particular magnetic element and the place of observation. For instance, at Batavia, south of the magnetic equator, the L variation of the declination increases in amplitude about ten-fold from the quietest to the most disturbed days of the year; for the declination at Greenwich the increase is only about two-fold.

(c) The external current system for L

The external current system which may produce the observed L-variation of the magnetic elements can be calculated by the same method as applied in the case of S_q. Figs. 15 and 16 depict the overhead current systems which would produce the L-variations at new moon at the equinoxes

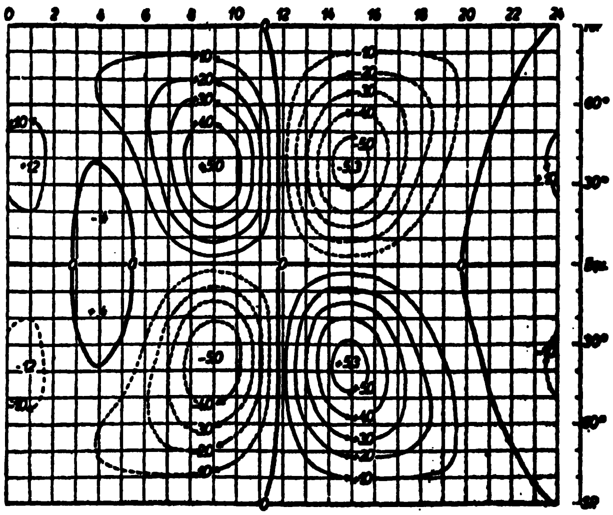


FIG. 15. Upper atmospheric current system at a height of about 100 km. which may produce the observed lunar diurnal variation of magnetic elements during the equinoxes. (After Chapman and Bartels.)

and at summer solstice respectively. The meridians refer to local lunar time; since the figure is for new moon, the sun is on the 12 hour meridian. The current lines are drawn 1,000 amperes apart. In the sunlit hemisphere at the equinoxes the currents flow in four closed circuits the intensity of each being about 5,300 amperes. In the night hemisphere the currents flow in two circuits one to the north and the other to the south of the equator—the intensity of each being about 1,200 amperes. This is in conformity with the greater range of L during daylight hours than during night. At the summer solstice the current flow is restricted to two main circuits in the sunlit hemisphere of which the stronger one carries about 11,000 amperes. This is more than double the current in the daylight equinoctial circuits. This is in accordance with the relatively large seasonal variation of L.

It is to be noted here that the intensity of the overhead current systems for S_q (Figs. 12, 13) is much greater than the corresponding current systems

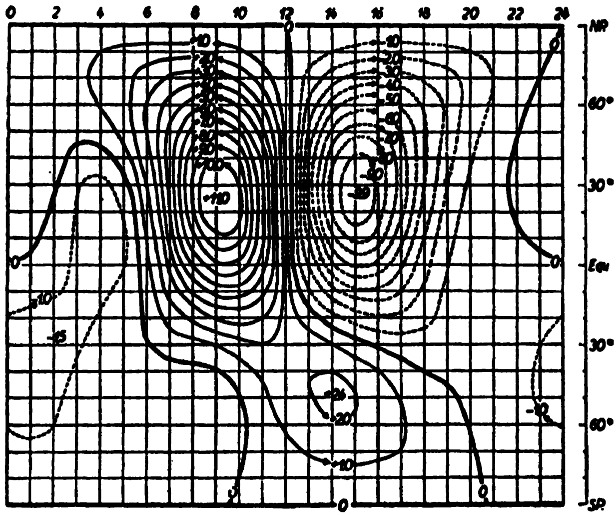


FIG. 16. Upper atmospheric current system at a height of about 100 km. which may produce the observed lunar diurnal variation of magnetic elements in June. (After Chapman and Bartels.)

for L (Figs. 15, 16). For instance, at the equinoxes, the ratio of the current intensity in each main circuit of S_q, to that of L is about 12 : 1.

5. THE DYNAMO THEORY OF S_q AND L

Various hypotheses have been proposed to explain the origin of the upper atmospheric current systems as would produce the S_q and L variations. Of these, the oldest, known as the *dynamo theory*, due to Balfour Stewart [24] and to Schuster [17] is now the most widely accepted one. (The theory was suggested for the S_q variation as the L variation was not known then.) This will now be discussed. Two other theories of S_q, one known as the *diamagnetic theory* due to Ross Gunn and the other as the *drift current theory* due to Chapman have also been proposed. Readers interested in these may refer to the original articles [25, 26] or to Chapter XXIII, Secs. 14 and 15, Vol. II of *Geomagnetism* by Chapman and Bartels.

Before entering into a discussion of the dynamo theory it would be helpful to have some idea of the electrical conductivity of the ionized layers of the upper atmosphere.

(i) *Conductivity of the upper atmospheric regions.*—It can be shown [27] that the conductivity due to a charged particle per unit volume of an ionized gas for an alternating electric field of angular frequency p is

$$\sigma = \frac{e^2}{m} \cdot \frac{\nu}{\nu^2 + p^2} \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad (3)$$

where e and m are the charge (in e.m.u.) and mass of the particle, and ν the frequency of collisions which the free charged particle experiences [see Appendix, Sec. 3(a)]. For a steady e.m.f. ($p = 0$) the formula reduces to

$$\sigma = \frac{e^2}{m\nu} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (4)$$

Assuming that there are electrons and ions of only one kind and that their respective number densities are n_e and n_i, we may write for the conductivity per unit volume,

$$\sigma = n_e \sigma_e + n_i \sigma_i \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (5)$$

where σ_e and σ_i refer to an electron and an ion respectively and are given by

$$\sigma_e = \frac{e^2}{m_e \nu_e} \text{ and } \sigma_i = \frac{e^2}{m_i \nu_i},$$

m_e and m_i being the mass of an electron and of an ion respectively.

If the ionized gas is under the influence of a magnetic field then the expression for σ differs for different directions with respect to the field. This is due to the gyration of the charged particles round the direction of the field. The conductivity in the direction of the magnetic field is unaffected by the presence of the field. The conductivity in a direction perpendicular to the magnetic field is given by

$$\sigma_{\perp} = \sigma \frac{\nu^2}{\nu^2 + p_H^2} = b\sigma, \text{ say,} \quad \ldots \quad \ldots \quad \ldots \quad (6)$$

where p_H ($= m/Fe$) is the angular gyro-frequency, and F is the intensity of the total magnetic field [27].

It is to be noted that in the direction of the magnetic field the conductivity increases as ν decreases and tends towards infinity as ν approaches 0. In directions perpendicular to the field, however, the transverse conductivity $\sigma_{\perp}$ attains its maximum limiting value when ν is equal to p_H and reduces to 0 as ν approaches 0 or tends towards infinity.

Both σ and b (Eq. (6)) vary with the nature of the charged particles so that the net transverse conductivity per cm.^{3} is given by

$$\sigma_{\perp} = \sum b\sigma. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (7)$$

If it is assumed as before that the medium contains electrons and ions of only one kind, we can write

$$\sigma_{\perp} = n_e b_e \sigma_e + n_i b_i \sigma_i \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (8)$$

where b_e and b_i refer to an electron and an ion respectively.

Tables III and IV show the conductivity of the atmosphere at various heights between 60 and 300 km. The number densities of electrons at levels 100, 150, 200, 250 km. given in the tables are based on measurements by radio method during a summer noon [see Chapter VI, Sec. 4(d)]. The values for regions in between are only probable values. It is assumed that the number of positive ions is equal to that of electrons at all heights.

The values of ν given in Table III are those which seem to be the most probable.

TABLE III

Conductivity in the direction of the magnetic field

Height (km.)	Number of electrons or ions per cm.^{3}	Collisional frequency of electrons. (ν_e)	Collisional frequency of ions. (ν_i)	σ_e	σ_i	$\sigma_\parallel$ $(= n_e\sigma_e + n_i\sigma_i)$ (e.m.u.)
60	$n_e = 5 \times 10^2$ $n_i = 5 \times 10^5$	2×10^7	$2{\cdot}5 \times 10^6$	$1{\cdot}4 \times 10^{-20}$	2×10^{-24}	8×10^{-18}
100	10^5	2×10^5	6×10^3	$1{\cdot}4 \times 10^{-18}$	$8{\cdot}5 \times 10^{-22}$	$1{\cdot}4 \times 10^{-17}$
150	$1{\cdot}5 \times 10^5$	5×10^4	$1{\cdot}6 \times 10^3$	$5{\cdot}6 \times 10^{-18}$	$3{\cdot}5 \times 10^{-21}$	$8{\cdot}4 \times 10^{-13}$
200	7×10^5	$1{\cdot}6 \times 10^4$	$5{\cdot}1 \times 10^2$	$1{\cdot}8 \times 10^{-17}$	$1{\cdot}1 \times 10^{-20}$	$1{\cdot}3 \times 10^{-11}$
250	10^6	$2{\cdot}5 \times 10^3$	$1{\cdot}1 \times 10^3$	$1{\cdot}1 \times 10^{-16}$	$1{\cdot}0 \times 10^{-19}$	$1{\cdot}1 \times 10^{-10}$
300	10^6	$1{\cdot}6 \times 10^3$	$6{\cdot}9 \times 10$	$1{\cdot}8 \times 10^{-16}$	$1{\cdot}5 \times 10^{-19}$	$1{\cdot}8 \times 10^{-10}$

TABLE IV

Conductivity transverse to the magnetic field

Height (km.)	(n_e, n_i)	b_e	b_i	$n_e b_e \sigma_e$	$n_i b_i \sigma_i$	$\sigma_\perp$ $(= n_e b_e \sigma_e + n_i b_i \sigma_i)$ (e.m.u.)
60	$n_e = 5 \times 10^2$ $n_i = 5 \times 10^5$	1	1	7×10^{-18}	1×10^{-18}	8×10^{-18}
100	10^5	$1{\cdot}4 \times 10^{-3}$	1	$2{\cdot}0 \times 10^{-16}$	$8{\cdot}5 \times 10^{-17}$	$2{\cdot}9 \times 10^{-16}$
150	$1{\cdot}5 \times 10^5$	$8{\cdot}9 \times 10^{-5}$	1	$7{\cdot}5 \times 10^{-17}$	$5{\cdot}3 \times 10^{-16}$	$6{\cdot}0 \times 10^{-16}$
200	7×10^5	$9{\cdot}1 \times 10^{-6}$	·96	$1{\cdot}1 \times 10^{-16}$	$7{\cdot}4 \times 10^{-15}$	$7{\cdot}5 \times 10^{-15}$
250	10^6	$2{\cdot}2 \times 10^{-7}$	·25	$2{\cdot}4 \times 10^{-17}$	$2{\cdot}5 \times 10^{-14}$	$2{\cdot}5 \times 10^{-14}$
300	10^6	$9{\cdot}1 \times 10^{-8}$	·12	$1{\cdot}6 \times 10^{-17}$	$1{\cdot}8 \times 10^{-14}$	$1{\cdot}8 \times 10^{-14}$

It is to be noted that for electrons the transverse conductivity $\sigma_\perp$ begins to differ from the longitudinal conductivity, $\sigma_\parallel$, at a height of about 100 km. whereas for ions the difference is negligible even up to 150 km.

(ii) *The dynamo theory*.—The dynamo theory as conceived by Balfour Stewart is that the S_q variation is caused by horizontal movements of the conducting upper atmosphere across the magnetic field of the earth. We may picture the conducting ionized region of the upper atmosphere as the armature of the atmospheric dynamo, the earth being its permanent field magnet and the atmosphere carrying the conducting layer the rotor.

If a conductor moves with velocity v in a direction at right angles to a magnetic field of intensity F then an e.m.f. E is induced in the conductor, given by $E = -vF$, per unit length of the conductor, where E, v, F are components along the positive directions of the X, Y, Z-axes of a left-handed co-ordinate system. In the case of the atmosphere, the motion is

caused by the tidal ebb and flow produced by the gravitational pulls of the sun and the moon. We have discussed in Chapter II the atmospheric pressure oscillations due to this tidal movement. The oscillations are found to be composed of a few sinusoidal pressure waves of which the most important is a solar semi-diurnal one of period 12 hours. The world distribution of this pressure wave may be represented, after Chapman [21], by the formula

$$\delta p = (AP_2^2 - BP_4^2)\sin[2(\lambda+t')+154°] \quad .. \quad .. \quad (9)$$

where δp is the change of pressure above or below the normal, P_2^2 and P_4^2 are the spherical harmonics which closely represent the pressure distribution, λ the longitude, t' the local time measured in degrees from the midnight meridian and A and B are constants [see Chapter II, Sec. 2(*b*)].

The main steps for calculating the e.m.f.'s generated are as follows:

The air velocity at a point in a wave-field, where the pressure change δp is small is given by a velocity potential ϕ, such that

$$\frac{d\phi}{dt} = -C^2\frac{\delta p}{p} \quad .. \quad .. \quad .. \quad .. \quad (10)$$

where C is the velocity of sound in the medium and p is the normal pressure.

Hence

$$\phi = -\frac{C^2}{p}\int \delta p\, dt.$$

Substituting for δp from Eq. (9) and integrating we obtain

$$\phi = \frac{NC^2}{4\pi p}[AP_2^2 - BP_4^2]\cos[2(\lambda+t')+154°], \quad .. \quad (11)$$

N being the number of seconds in a day. The northward and eastward components of the air velocity are therefore given by

$$\left.\begin{aligned} u_N &= \frac{d\phi}{a\,d\theta} \\ u_E &= \frac{d\phi}{a\sin\theta\, d\lambda} \end{aligned}\right\} \quad .. \quad .. \quad .. \quad (12)$$

where θ is the colatitude and a the earth's radius. The velocities being known from the above equations it is easy to calculate the induced electromotive forces due to the motion of the charged particles across the vertical component of the earth's magnetic field Z. Thus

$$\left.\begin{aligned} \text{Eastward e.m.f.} &= -u_N \times Z \\ \text{Northward e.m.f.} &= u_E \times Z \end{aligned}\right\}. \quad .. \quad .. \quad .. \quad .. \quad (13)$$

The air-velocities as calculated [28] from Eqs. (12) are shown in Fig. 17. The corresponding induced electro-motive forces as obtained from Eqs. (13), assuming the value of Z to be constant along each parallel of latitude, are represented in Fig. 18. It will be seen that at the equator where the value of the vertical component of the magnetic field is zero, the

induced e.m.f. is also zero. The current in this region is driven by the e.m.f.'s induced in higher latitudes.

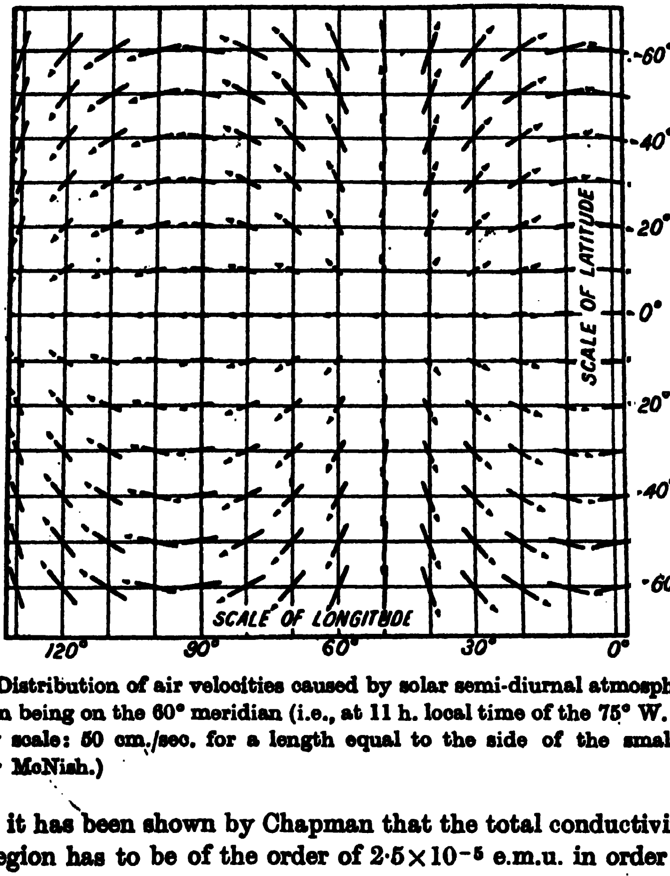


FIG. 17. Distribution of air velocities caused by solar semi-diurnal atmospheric tides, the sun being on the 60° meridian (i.e., at 11 h. local time of the 75° W. meridian) Arrow scale: 50 cm./sec. for a length equal to the side of the small squares. (After McNish.)

Now it has been shown by Chapman that the total conductivity of the ionized region has to be of the order of $2{\cdot}5\times10^{-5}$ e.m.u. in order that the current strength may be that necessary for producing the observed S_q variation. The total conductivity $\int \sigma\, dh$ between 60 and 300 km. may also be calculated with the help of Table IV. (We use Table IV rather than Table III because the currents have to flow over a large portion of the circuit in directions transverse to the magnetic field.) The total conductivity thus obtained is of the order of 10^{-7} e.m.u.

There is thus a discrepancy of two orders between the total conductivity obtained from radio measurements and that demanded by the dynamo theory. This has been regarded as one of the stumbling blocks of the dynamo theory. The following considerations show, however, that this difficulty is to a large extent more apparent than real.

In calculating the e.m.f.'s due to tidal motions the velocities of the air particles have been deduced from the observed amplitudes of barometric oscillations near the ground. But as shown in Chapter II the mode of atmospheric oscillations is such that the velocity of tidal motion above 30 km. (height of the nodal surface) is about 200 times the velocity below

this height. The e.m.f. developed should therefore be 200 times that calculated above. The total conductivity necessary to produce the required intensity of S_q current system need thus be only 1/200th of that mentioned above ($2{\cdot}5 \times 10^{-5}$ e.m.u.).

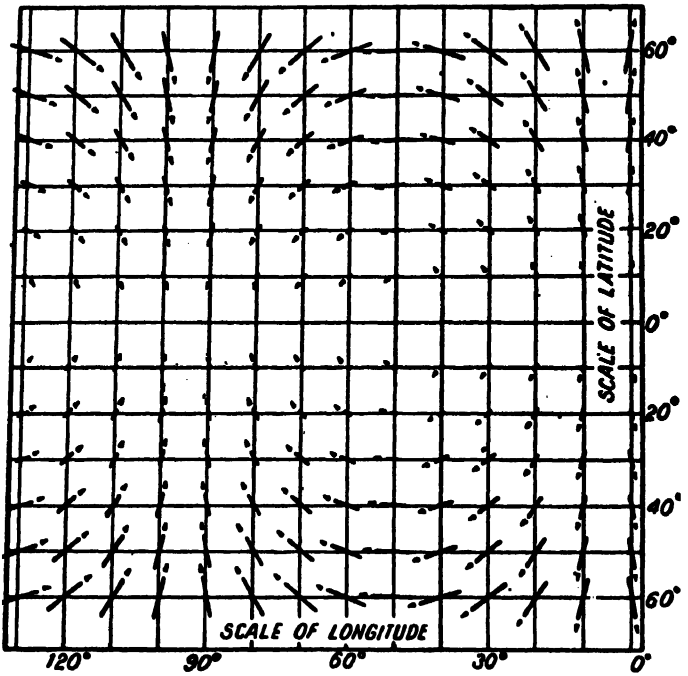


FIG. 18. Distribution of electromotive forces caused by the air motion depicted in Fig. 17 assuming the earth's magnetic field to be that due to a dipole at the centre with axis along the geographical axis of the earth. Arrow scale: 0·023 volt/km. for a length equal to the side of the small squares. (After McNish.)

Another explanation of the difficulty that is sometimes suggested may also be mentioned. It is possible that there is copious production of negative ions on account of the strong electron affinity of the O-atoms (and also, to a lesser extent, of the O_2-molecules). Hence, contrary to the assumption made above, the number density of the ions (positive and negative together) might, instead of being the same as that of the electrons, be several orders higher. The conductivity of the current carrying regions would also be correspondingly higher and may thus be adequate to explain the S_q variation.

(iii) *Some unexplained features of S_q and L variations*.—The dynamo theory in its simple form as discussed above cannot explain all the details of the S_q and L variations. We mention below some of the difficulties.

It has already been indicated that the S_q and the L current systems undergo large seasonal changes; they are more intense in summer than in the equinoxes. But, while the summer/equinox ratio for the L current

systems is 2·6, that for the S_q system is only about half as much. (This anomaly is as yet unexplained.)

Both S_q and L current systems increase from sunspot minimum to sunspot maximum of the solar cycle. But, contrary to what may be expected from the nature of the seasonal variations, the increase in S_q is greater than the increase in L. Thus, while S_q increases by about 50%, L increases by only about 20%. (This may necessitate the supposition that the S_q and L current systems lie in different ionospheric layers with some phase difference between their motions.)

At Huancayo (South America) situated on the magnetic equator and between the north and south current foci, the range of diurnal variation is found to be twice as large as at other places. (See Fig. 14.) This has been tentatively explained by McNish as due to the obliquity of the geomagnetic axis. But, according to Egedal [29] this increase in the range of H is confined in a narrow zone, near the magnetic equator, and may be produced by a varying electric current flowing in a very narrow strip at a height of about 100 km. above the magnetic equator.

The intensity and the foci of the S_q current systems are found to vary quite appreciably from day to day as has been studied in considerable detail by Hasegawa [15]. The latitudes of the foci may alter for example, by as much as 15° producing different variations along a line of latitude. (Such variations may be due to variations of conductivity and/or of wind velocity in the ionospheric layer where the current system lies.)

6. MAGNETIC DISTURBANCES—THE S_D, D_{st} AND D_i VARIATIONS

As mentioned before, the disturbed parts of the magnetic elements might be grouped into two main categories. There is one part S_D which varies with the local time and there is another part D_{st} which is symmetrical about the earth's axis and occurs simultaneously over the whole world.

In the present section we shall first deal with the nature of the S_D and the D_{st} variations and then discuss the upper atmospheric current systems which may produce the observed variations.

(a) The S_D field

It has been found that the nature of the S_D variations depends to a marked degree on the latitude of the place of observation. An inspection of the horizontal force S_D curves given in Fig. 6, columns b and c, shows that in the middle and moderately high latitudes they differ from the S_q variation in the following respects:

(i) Unlike S_q the amplitude of the S_D variation is approximately the same during day as during night.

(ii) The phase of the variation is reversed somewhere between 51° and 56° latitude.

(iii) The amplitude decreases from equator up to this latitude and then increases.

For regions of higher latitude the S_D variation differs markedly from that in regions of lower latitude, i.e., between latitudes $\pm 60°$. The form of the S_D field in the polar regions is closely associated with the auroral belts.

A remarkable feature of S_D in high latitudes is that it rapidly begins to exceed D_{st} as the auroral zone is approached. The variations in H and V of the S_D part of the field, as one proceeds from lower to higher latitudes, are depicted in Figs. 6 and 8, column b, and in Fig. 19 which represents the annual mean 'all-days minus quiet-days' S_D variation of vertical force in polar regions. It will be noticed from these figures that unlike the S_D curve for H, the S_D curve for the vertical force does not change sign at 56°. The amplitude of the V-curve goes on steadily increasing from the equator to this latitude. At the magnetic latitude of 64° (Nova Zembla), i.e., near the isochasm of maximum auroral frequency the amplitude of V decreases considerably. (Lines of equal frequency of occurrence of auroras are called *isochasms*.) Beyond this latitude, as one proceeds inside the auroral zone, the vertical force S_D again increases in amplitude but with the phase reversed.

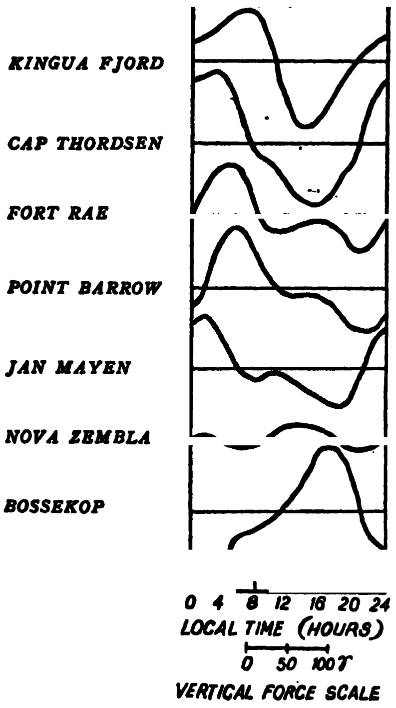


FIG. 19. Annual mean S_D variation (all-days minus quiet-days) of vertical force in polar regions. The magnetic latitudes of the stations are: Kingua Fjord 79°, Cap Thordsen 74·5°, Fort Rae 69°, Point Barrow 68·6°, Jan Mayen 73·4°, Nova Zembla 64°, Bossekop 67°. (After Chapman.)

The horizontal force S_D also undergoes marked changes as one approaches the auroral zone. Unlike the change in the vertical force S_D there is not merely a change in phase, but also a striking change of type. This is most conveniently shown by the horizontal force vector diagram for S_D. These curves are drawn in the following manner. A vector OP is drawn from a fixed origin O, its length representing the magnitude of the S_D part of H and its direction the declination. As this length and the declination vary the point P, plotted over 24 hours describes a closed curve the non-periodic variation having been allowed for. Fig. 20 depicts such curves which are derived from all-days minus quiet-day means. Between latitudes $\pm 60°$

these curves are roughly of oval form. At Sitka (mag. lat. 60°) the curve has still some resemblance to a rough oval form; but at Sodankyla

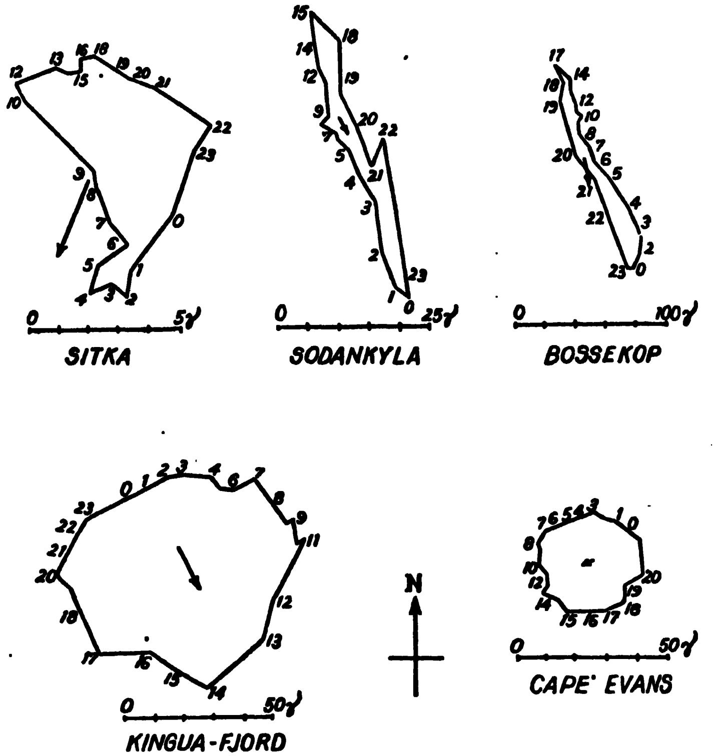


FIG. 20. Horizontal force S_D vector diagrams for high latitudes. The magnetic latitudes (north) of the stations are: Sitka 60°, Sodankyla 64°, Bossekop 67°, Kingua Fjord 79°, Cape Evans 83°. (After Chapman.)

and Bossekop (mag. lats. 64° and 67°) the curves are very narrow and elongated. These two stations are nearly under the zone of maximum auroral intensity. Closer inspection shows that the direction of elongation is nearly normal to the direction parallel to the auroral zone. At Kingua Fjord and Cape Evans (mag. lats. 79° and 83°), stations which are well beyond the auroral belt to the north, the curves regain their oval forms and are, in fact, approximately circular. It should be noted that the vector diagrams for the different stations are drawn to different scales. The scale is so chosen that the figures are of similar size so that the change of form may be easily studied.

The reversal of phase of the vertical force S_D variation and of the sudden change of phase and type of the horizontal force S_D variation within the narrow region of the auroral belt, on either side of which the range of variation itself is so large, constitute perhaps the most striking feature of all the average characteristics of world-wide terrestrial magnetic disturbances.

(*b*) The D_{st} field

As indicated before there is a part of the D field which is not dependent on local time but which is of a world-wide character affecting both the northern and southern hemispheres simultaneously. The method by which the characteristic features of these world-wide variations—the so-called magnetic storms—can be separated from other irregular and local time variations has already been indicated [Sec. 2(*b*)]. According to their intensities the storms are classified as great (200γ), active (100γ) and moderate (50γ). (The values enclosed within the brackets indicate the maximum range of variation.) Typical storm-time variations of the magnetic elements for storms of *moderate* intensity as obtained from records of 11 observatories for a set of 40 storms, are shown in Fig. 9. The three sets of diagrams are for three sets of stations grouped in different magnetic latitudes as indicated in the figure. The characteristic features of these variations are as follows:

The horizontal force variation commences suddenly. (Usually denoted by *S.D.-sudden commencement.*) It at first increases above the normal undisturbed value, remains so for a few hours and then decreases rapidly, attaining a minimum value which is generally greater than the initial increase. It then takes several days for the depressed value of H to regain its normal value. The recovery is at first rapid and then slows down considerably. The increase is called the *initial phase* of the storm. The decrease up to the point at which the rate of recovery begins to slow down is called the *main phase.* The period of slow recovery is called the *last phase* or the phase of recovery of the storm.

The change in the vertical force is much smaller than the change in the horizontal force and occurs in the opposite sense; the average change in declination is also small particularly in lower latitudes. Further, the variation of H decreases with increasing latitude.

These considerations show that the storm time field over the range of latitudes depicted in the figure is of a very simple form. It is symmetrical about the earth's magnetic axis; it has one direction during the initial phase and this direction is reversed during the main and the last phases of the storm.

(*c*) Current systems for S_D and D_{st}

Harmonic analysis of the D_{st} characteristic variations shows that the main part of the storm time field is due to causes above earth's surface [30]. No harmonic analysis of the S_D field has been attempted. But it is reasonable to suppose that this field also is due principally to causes external to the earth. It is therefore possible to determine the nature of the current systems which, independent of any theory as to their origin, should flow overhead to produce the observed S_D and D_{st} variations. (In Chapter IX current theories of the origin of magnetic storms will be discussed.)

We will first describe the somewhat idealized current systems, as deduced by Chapman [30a], for the average S_D and D_{st} field of the set of 40 storms of moderate intensity mentioned above.

In Fig. 21, (*a*) and (*b*) represent the current system for the S_D variation as viewed from the sun and from above the north pole respectively. The central vertical line in (*a*) represents midday; the circle in thick line round the N-pole in (*b*) represents the flow of current along the auroral belt. It may be noted that the currents along the periphery of the two halves of the circle are in opposite directions so that the circuits in each half are completed by the parallel lines drawn inside the circle. A part of the current

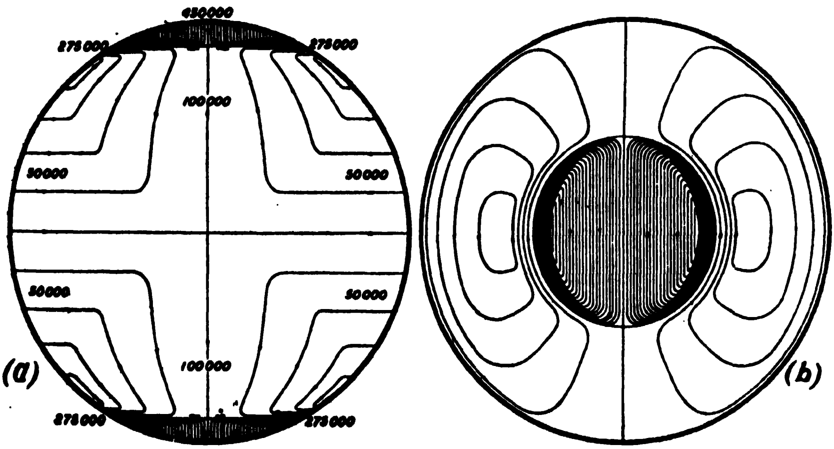


FIG. 21. Idealized atmospheric current systems producing S_D variations as deduced from forty magnetic storms of moderate intensity. (After Chapman.)
(*a*) View from the sun. (*b*) View from above the north pole.

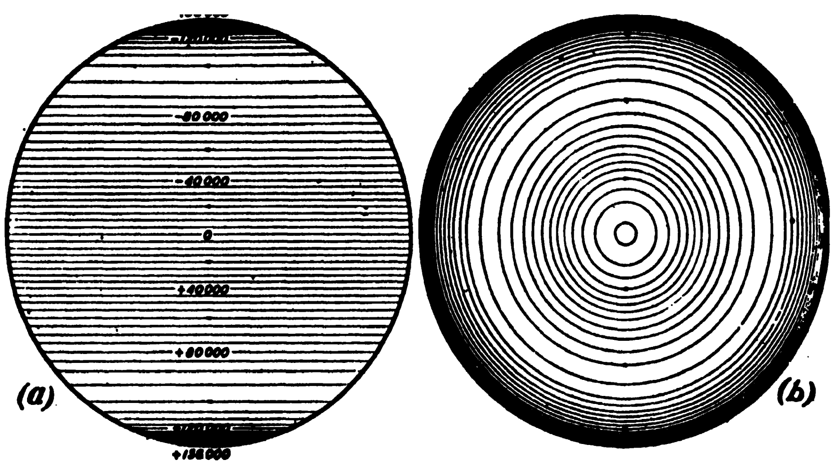


FIG. 22. Idealized atmospheric current systems producing D_{st} variations as deduced from forty magnetic storms of moderate intensity. (After Chapman.)
(*a*) View from the sun. (*b*) View from above the north pole.

may also be completed by flowing round the earth; this is shown by the central vertical line.

In Fig. 22, (*a*) and (*b*) represent the current system for the D_{st} variation as viewed from the sun and from above the north pole respectively.

In Fig. 23, (*a*) and (*b*) represent the combined S_D and D_{st} current systems viewed as above.

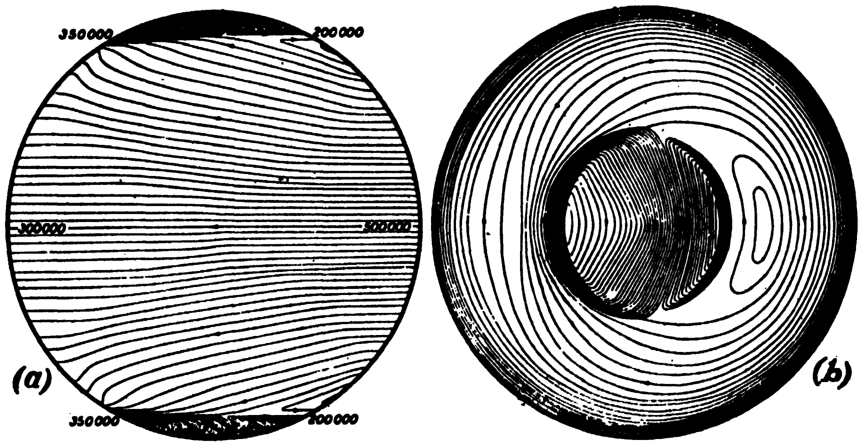


FIG. 23. Idealized combined atmospheric current systems producing both S_D and D_{st} variations (as depicted in Figs. 21 and 22) during magnetic storms of moderate intensity. (After Chapman.)

(*a*) View from the sun. (*b*) View from above the north pole.

It is to be noted that the current systems are fixed as it were with reference to the sun. The earth rotates with reference to this current system, so that a point on the surface of the earth at a particular latitude is brought under different portions of the current flowing in that latitude.

Consider now the S_D current systems as shown in Fig. 21. The current system would produce the following daily characteristic variations of the three components X, W and V on the surface of the earth.

At the equator and on both sides up to the latitudes where the two closed current systems shrink to points, the north horizontal force (X) variation will have its maximum at 6 hour and minimum at 18 hour. At the latitude noted above the phase of X variation will be reversed. From here up to the auroral belts this reversal will be maintained.

On the two sides of the equator the variation of magnetic west component (W) will have opposite phases. Unlike X it will not suffer a phase reversal between the equator and the auroral belt.

From equator up to the auroral zone the vertical force (V) diurnal variation will have the same phase; it will have opposite phases on the two sides of the equator. In northern latitudes the downward force will be maximum at 18 hour and minimum at 6 hour; the daily range of variation will increase from zero at the equator to a maximum at a point a little to the south of the auroral belt.

An inspection of curves in columns *b* and *c* of Figs. 6, 7 and 8 shows that all these variations are in general agreement with the observed data.

The current system depicted for the polar cap within the auroral zone will also yield the observed magnetic variations as given in Figs. 19 and 20. We can infer from the current system that, as the auroral current belts are approached the horizontal force variation will become more and more directly transverse to the zone. W-variation will decrease in range and will vanish under the current belt while the X-variation will remain the same in phase (minimum in the morning and maximum in the afternoon) and increase greatly in magnitude. The vector diagram of H will thus become rectilinear with its direction normal to the current belt and with its north and south elongations at 18 hour and 6 hour respectively. As we cross the belt the range of X will decrease but it will preserve its phase. W-variation will reappear but with reversed phase. H-vector-diagram will thus again become oval but will be described in a sense reverse (counter-clockwise) to that outside the auroral zone. Further, within the zone X-variation will vanish and will reappear with reversed phase while the W-variation will have the same phase everywhere within the zone. The sense of description of the H-vector-diagram will thus be again reversed. There will therefore be three reversals in all of the horizontal force vector-diagram as one proceeds northward from the equator at latitudes of about 55°, 65° and 75°, the intermediate latitude being that of the auroral current belt. From the fact that the H-vector-diagram within the auroral zone near the pole is nearly circular we may regard the H-variation as due to the rotation of the earth within a magnetic field which near the poles is roughly uniform and horizontal. The current system producing such a field is also necessarily a sheet of parallel uniform current.

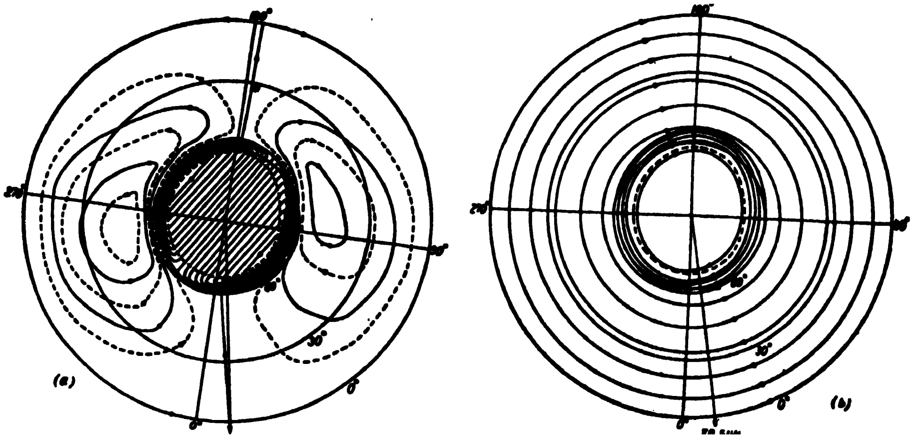
With regard to the vertical force we easily see that as the auroral belts are approached the phase of V-variation will remain unaltered up to the centre of the current belt; beyond this the phase will be reversed. The range of V will have a maximum at a latitude slightly lower than that of auroral belt and will decrease rapidly to zero value at stations directly under the belt. As we proceed further into the zone the range of V-variation will rapidly rise to another maximum a little to the north of the belt and from there will gradually decrease to zero at the pole that is the centre of the auroral zone.

The current system for D_{st} shown in Fig. 22, (*a*) and (*b*) is quite simple being everywhere parallel to the circles of latitude, directed east to west. It is easy to see that the magnetic field associated with such a current system will be directed southwards everywhere and will be most intense under the two current belts. These inferences agree with the observed facts. The vertical force will vanish at the equator and will be large and of opposite signs near to and on opposite sides of the current belts.

It is to be remarked that these inferences with regard to the variation of the magnetic elements are only in general accord with the observed facts. There are discrepancies, but this is not surprising, since, in the

current diagrams drawn the magnetic axis of the earth has been taken as coinciding with its rotational axis. The inclination between the two is, however, not inconsiderable, being about 11·5°.

It will, therefore, be interesting to compare the idealized current systems of Figs. 21, 22 and 23 (average of 40 storms) with the current systems for an individual storm. For this, we reproduce in Fig. 24 the computed current systems for an intense magnetic storm during the Polar Year 1932-33. The computations are by Vestine [29a] and the figures depict the overhead currents as might have flowed at height of 150 km. on May 1, 1933, at 16 h. to produce the observed S_D and D_{st} fields, at that hour, of an intense magnetic storm which commenced the previous day. (The original diagrams also showed the geographical distributions of the mean hourly disturbance vectors as recorded at the various observing stations. These have been omitted from the figures.) In computing the currents allowance was made for the effects of the induced earth currents. An inspection of the figures at once shows that the S_D (Fig. 24a), the D_{st} (Fig. 24b) and the



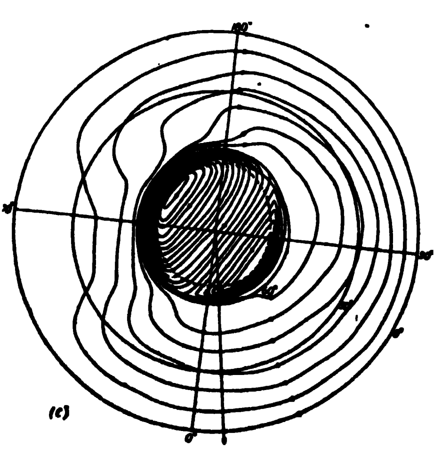


FIG. 24. (a) S_D, (b) D_{st} and (c) the combined current systems for an individual storm of large intensity as compared to the idealized current systems for storms of moderate intensity in Figs. 21, 22 and 23. The storm occurred on April 30-May 1, 1933, and the current systems are those as existed at 16 h. on May 1, 1933, during the main phase. The computations are made on the assumption that the currents flowed at a height of 150 km. 100,000 amperes are flowing between successive full-drawn lines. (After Vestine).

combined current systems (Fig. 24c) closely resemble the respective idealized current distributions in Figs. 21, 22 and 23 respectively. The intensity of the currents in the former is, however, about four times that in the latter. This means that the storm was about four times as intense as the average storm in Chapman's computation.

The computations of the current systems as made above are from general considerations of the observed variations of the magnetic elements without taking recourse to detailed harmonic analysis. This has been done by Slaucitajs and McNish [30]. The current systems obtained by these authors from spherical harmonic analysis of magnetic storms data also show three regions of maximal intensity, one near the equator and two others near the auroral zones, in fair agreement with the current system of Chapman described above.

It should be mentioned that besides the current systems which produce the regular features of the magnetic storms, there are also sporadic electric currents which produce many irregular features. We mention here two instances [31]:

There are local vortices of electricity known as Schmidt's wandering vortices about the size of the continent of Europe. In their form and motion these vortices may be compared to the cyclones and anti-cyclones of the troposphere.

Another irregular feature has been discovered by Sangster. Sangster made a detailed study of H-, D- and Z-curves of a large number of magnetic storms using the Greenwich records and found that the total force vector executes at some periods a clockwise and other periods an anti-clockwise rotation. This phenomenon seems to have some connection with the behaviour of current systems generated in a conducting sheet as studied by Price. Price showed that if the sheet is of uniform conductivity, the current would die away where it was. If, however, the conductivity is non-uniform, then the current system, during the process of decay, would move towards the parts which are more conducting.

(*d*) The energy of magnetic storms

During a magnetic storm the total energy of the magnetic field in and around the earth is increased. For a medium of permeability unity, the energy density E in a magnetic field F is given by

$$E = \frac{F^2}{8\pi}.$$

If F is changed by ΔF, the corresponding energy change is given by

$$\Delta E = \frac{F \Delta F}{4\pi}.$$

During the initial phase of the storm, ΔE rises. During the decay of the disturbance field the energy is withdrawn and dissipated in the form of heat produced by the atmospheric current systems and the earth currents

Apart from this, during the production and increase of the current systems, and during the period of gradual relaxation before free decay, there is continual loss of energy in overcoming the electrical resistances. During this phase of the storm the electromotive forces supply the heat loss and also the energy of the magnetic field.

According to Chapman [32] the amounts of energy lost as heat during the rise and the decay of the storm field are nearly equal. The energy loss during free decay is approximately equal to ΔE_0, the maximum excess energy of the magnetic field during the storm. Thus, the total energy supplied by the external e.m.f. during the course of the storm is $2\Delta E_0$. The maximum excess energy ΔE_0 is thus a measure of the energy of a magnetic storm. Chapman has shown that ΔE_0 is nearly proportional to ΔH_0, the maximum mean diminution of horizontal intensity on the equator.

A storm for which ΔH_0 is 100γ, which may be taken as a standard storm of fairly large magnitude, the total energy loss has been estimated to be of the order 10^{23} ergs. This is supplied during the initial phase of the storm which is of the order 15 hours. Hence the rate of supply of energy is 2×10^{18} ergs. sec.$^{-1}$ This is but a small fraction of the energy from the sun intercepted by the earth, namely, 2×10^{24} ergs. sec.$^{-1}$

(e) The irregular disturbances

As already mentioned [Sec. 2(b)], there remains a part of the disturbance D_i field, after S_D and D_{st} have been separated out, which is wholly irregular. Being related neither to the solar time nor to the storm-time, these irregularities are smoothed out in the process of determining S_D and D_{st}.

But, though D_i shows almost infinite variety in detail, it has been found to possess on the average certain characteristic features. Thus, D_i is markedly dependent on latitude and also on local time. This means that its intensity has a well-defined geographical distribution and that like S_q and S_D, at any particular season, this distribution is fixed as viewed from the sun.

We describe below after Chapman [33] these characteristics in some detail.

D_i is of small intensity at low latitudes. With the increase of the latitude the intensity increases, the increase being very rapid as the auroral zone is approached. After a pronounced maximum under or nearly under the auroral zone the intensity decreases as one proceeds inside the zone. The decrease has a seasonal variation. In summer it decreases to about one-half of the maximum (as under the auroral zone) and in winter to about one-fifth. The minimum winter value of intensity is, however, several times larger than the equator value (where there is little seasonal variation).

The form of the diurnal (local time) variation of D_i changes with the latitude. In the latitudes between the auroral zones (i.e. say between

± 65° mag. lat.) the form is simple with maximum in the evening. This form of the diurnal variation does not change much either with the season or with the general intensity of the disturbance, except that the hour of maximum gets later, being moved from 21 h. at 55° to midnight at 70°. Between 70° and 80° magnetic latitudes (i.e. well within the auroral zone) the character of the diurnal variation changes markedly. It now depends both on the season and also on the general intensity of the *D*-field.

Nikolsky has made a close examination of the D_i field in high magnetic latitudes from the data of 32 stations (in particular, of the Tikhaya station 71·5° N.) and is of opinion that too little attention has hitherto been paid to the study of this irregular field [34]. According to Nikolsky all the magnetic storm phenomena are caused *directly* by corpuscular streams for the sun which produce only accidental, discrete and short period disturbed magnetic fields. He found a high correlation between the accidental, irregular D_i field (not its average values) during storm time and the corresponding absolute value of H. Further, if the disturbed days data are taken and from them individual quiet hours are selected, then it is found that the values of H for these quiet hours are exactly the same as those on the quietest days. From this Nikolsky concludes that the persistent field changes during magnetic storms as determined by averaging the disturbed days data (in which process the irregular D_i fields are smoothed out) do not really exist. The S_D and D_{st} current systems as pictured in Figs. 21 and 22, and to which the persistent field changes are ascribed, are thus merely statistical results and fictitious.

Nikolsky has further found (from Sodankyla records for the Polar Year 1932-33) the presence of two maxima in the diurnal variation of magnetic disturbances, one in the morning and the other in the evening. According to him these two maxima depend on different laws, and, as such, are due to different causes, e.g. charged particles of different signs—positive and negative.

It is to be mentioned that in the polar regions the magnetic activity has a distinct diurnal variation being maximum at about zero hour local magnetic time [37]. (See Fig. 25.)

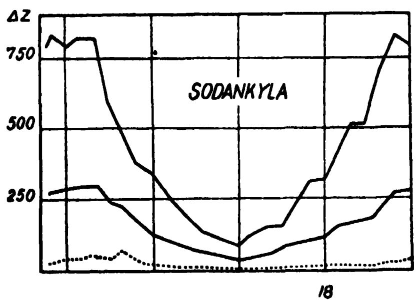


FIG. 25. Illustrating diurnal variation of magnetic activity (expressed by ΔZ-values) near the polar regions. The upper curve is for *disturbed* days; the middle curve for *all* days and the lower curve is for *quiet* days. The abscissa is local geomagnetic time. (After Sucksdorff.)

(*f*) Estimation of the height of the current belts and the intensity of the currents in the auroral zones

It is possible to estimate the height at which the current carrying belts in the auroral zones (thick lines in Fig. 21) are situated by making simultaneous observations on the resultant S_D variation at two stations in the immediate vicinity of these belts. The method of doing this is quite simple and straight-forward. Two stations are chosen very near the auroral belt but at different perpendicular distances from the same. The direction of the resultant S_D field, due to the approximately linear overhead current belt is found out from the H and V variations at the two stations. Lines drawn perpendicular to the resultant S_D vectors (in the vertical plane) will obviously pass through the current belt. The height at which the two lines from the two stations cross each other will give the height of the current belt. Birkeland [35] estimated the height in this manner and found it to vary from 150 km. to 600 km. Goldie's estimate

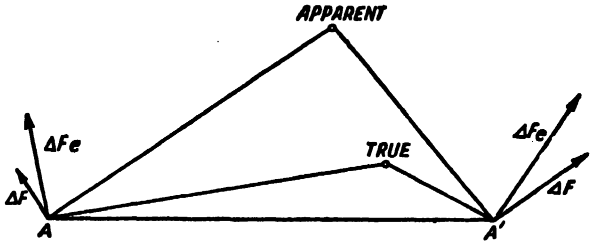


FIG. 26. Estimation of height of current belt in the auroral zone. A, A'—two stations on opposite sides of the belt. F—total disturbance field, F_e—external part of the disturbance field.

is between 290 and 370 km. It has been pointed out by Chapman [36] that these estimates suffer from considerable uncertainty owing to the fact that no account has been taken of the effect of the secondary internal current systems induced within the earth (*vide infra*). It seems likely that the H-variations due to external and internal current systems will usually be similar in direction or sign but the V-variations will have opposite signs. As a consequence the inclination of the observed resultant S_D field is less upwardly directed than the true resultant S_D field due to the external current belt alone. The estimated height, if the effect of the internal current system is not allowed for, may, therefore, greatly exceed the true height. This is illustrated in Fig. 26. It is necessary that further theoretical study of the nature of the induced earth currents be made in the polar regions for making more accurate determination of the height of the auroral current system.

According to estimates of Sucksdorff, based on the 1932-33 Polar Year data of a number of high latitude stations, the average intensity of the belt-current in the auroral zone is 120,000 amperes, varying between less than 100,000 amperes on quiet days in winter to about a million amperes on disturbed days in summer [37]. The average position of the current

belts is near the auroral zones, but it moves out during the disturbed periods and falls slightly in height. Of the two branches of the current belt (the two thick-lined portions in Fig. 21(*b*)), the *W*-current branch predominates chiefly in the first part of the day and the *E*-current branch in the latter part. It appears that the afternoon current towards east is more regular and permanent than the other branch.

Besides the nearly horizontal auroral belt current, Sucksdorff reports the presence of nearly vertical strong current systems in the vicinity of the magnetic axis poles. This current system has also two parts, one forenoon and another afternoon. The horizontal component of the forenoon part is eastward and that of the afternoon part westward. The current is very intense, the average intensity being 350,000 amperes, or, nearly double the intensity of the maximum afternoon current of the auroral zone. During high magnetic activity both the intensity and the height of the currents increase. The direction of the forenoon east-current, at a distance of about 1° from the axis pole, is away from the earth; that of the afternoon west-current, close to the axis pole, is towards the earth (i.e. if the currents are considered as positive). It is to be noted that for both the vertical polar current and the horizontal auroral belt current the maximum of intensity and height of the afternoon current occurs at the same local time.

(*g*) Earth-currents

Before closing this section we shall refer briefly to the nature and method of studying the so-called earth-currents, or, currents flowing in the crust of the earth presumably induced by the overhead current systems.

For studying the earth-currents potential differences between widely separated points in the earth's crust are measured [38, 39, 40, 41]. Three points forming a right angle, the arms of which are directed north-south and east-west, or better, four points such that the line joining one pair is at right angles to that joining the other pair, are selected. Large metallic electrodes are buried at these points and the corresponding pairs are connected by means of well-insulated underground cables or by overhead lines. Sensitive instruments are inserted at suitable points in the lines for reading the voltage difference between the electrodes which is of the order of a few milli-volts per kilometre distance. The distance between the electrodes forming a pair varies widely from less than a kilometre to 200 kilometres. Since the electrodes, as well as the soils in which they are embedded, are never identical physically and chemically, a certain potential difference always exists between them. This potential difference does not remain constant but varies considerably with time due to various causes. Allowance for such contact potentials has therefore to be made in order to obtain the true value of the potential gradient at the time of measurement. It is also necessary that the contact resistance at the electrodes be very small compared to the resistance in the recording circuit and that it does not vary with time. The usual practice is to employ a grid of lead wire, of area 25 to 40 sq. metre buried in moist clay [42, 43].

In certain measurements on earth-currents made near Bombay, India, Banerji overcame the polarization difficulties by making the electrodes (the distance between them being short—only 250 m.) neutral with respect to the soil, each electrode being a combination of electro-positive and electro-negative metals [41]. The actual composition of the electrodes however varied with the soil and had to be found by trial. The neutrality did not maintain itself indefinitely, but could be compensated by suitable device.

The earth-current variations, according to the records, may be broadly classified under two heads: (i) irregular fluctuations or earth-current storms, and (ii) regular or periodic variations, such as diurnal and seasonal variations.

The earth-current storms have many features in common with the magnetic disturbances. They occur simultaneously with the magnetic disturbances and, like the latter, vary widely both in type and in intensity. World-wide storms commence suddenly and simultaneously all over the globe. Local disturbances occur with great intensity at high latitudes. And, like the magnetic activity, the earth-current activity shows a variation of eleven-year period corresponding to the sunspot cycle. It will be noticed that the earth-current records are closely parallel to and are of the same nature as the fluctuations in the magnetic records.

The diurnal variations of the earth-currents may be obtained by plotting the average hourly departures from the daily mean value of the potential gradient [41, 44, 45, 46, 47].

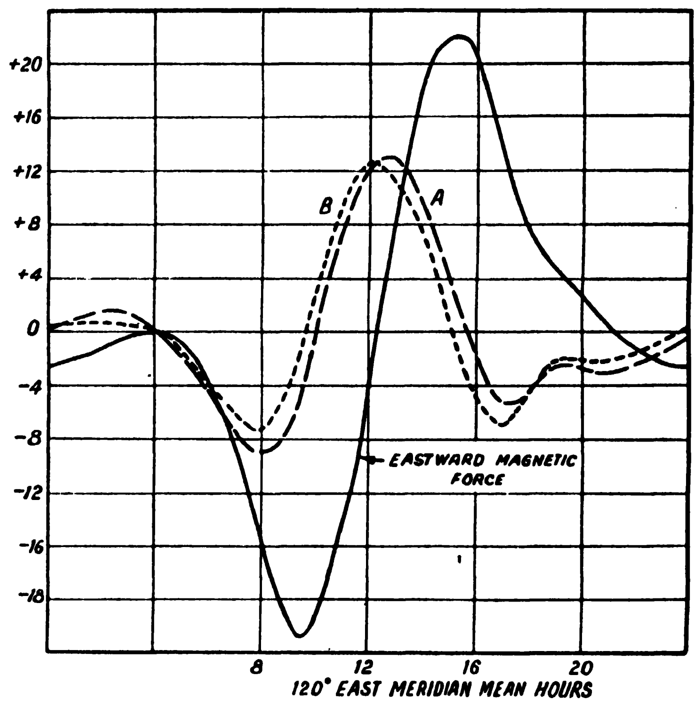


FIG. 27. Diurnal variation of rate of change of magnetic force and of earth currents. (After Rooney.)

In Fig. 27, the diurnal variation of the northward component of the earth-current (curve A) and the time derivative of the eastward component of the horizontal magnetic force (curve B) for Watheroo (Lat. 30° S.) are shown. The variation of the eastward magnetic force is depicted by the full-line curve. Close parallelism between the curves A and B at once suggests a causal relation between the two. It is, in fact, quite reasonable to conclude that the earth-currents—at least their major fractions—are merely the induced currents due to the fluctuations in the terrestrial magnetic force.

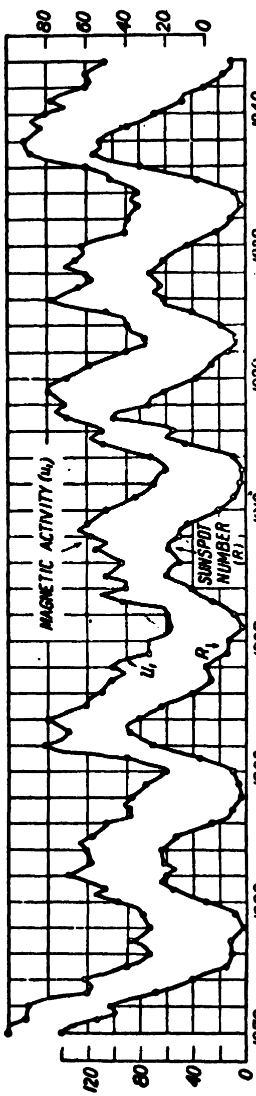


FIG. 28. Illustrating the remarkable correlation between the magnetic activity u_1 and sunspot number; the upper curve represents the annual means of u_1 while the lower curve represents those of the sunspot number. (After Bartels.)

There may, of course, be other factors contributing to the earth-currents. There is, for instance, a systematic phase difference between curves A and B. Again, at stations in middle latitudes where the variation of the eastward component of the earth-current is quite pronounced, the diurnal variation is found to resemble the variation curve of the northward magnetic component (reversed) and not its (the latter's) time rate of change as it ought to, if the current systems were due to induction effect. The existence of such variations shows that the origin of the totality of the earth-currents is more complex than that suggested by the marked similarity of the curves A and B.

7. MAGNETIC DISTURBANCES AND SOME SOLAR PHENOMENA

We have in Secs. 3-4 referred to the dependence of the S_q and L variations on the 11-year solar cycle. In the present section we shall discuss some remarkable correlations which exist between certain solar phenomena on the one hand and terrestrial magnetic disturbances on the other. The existence of such correlations shows that the causes of the S_D and the D_{st} current systems, as depicted in Figs. 21 and 22 also lie in the sun. The probable nature of such causes will be discussed in Chapter IX.

Examination of data collected for many years shows that the terrestrial magnetic disturbance is more frequent and on the average more intense at the epoch of sunspot maximum than at sunspot minimum. The amplitude of the D-variation thus undergoes a cyclic change of the same period as that

of the sunspots, namely, approximately 11 years. This is illustrated in a striking manner in Fig. 28 showing the variation of mean sunspot numbers and the annual mean magnetic activity u_1 over the long period 1870–1944; the parallelism between the two curves is indeed remarkable. The magnetic activity (u_1) is, as already explained, a measure of the average annual value of D_{st} [Sec. 2(a)]. The same phenomenon is also depicted in another manner in Fig. 29. The upper part of the figure shows the monthly means of (u) for the period 1900–1930, while the lower part shows the corresponding sunspot numbers [48, 49, 50, 51].

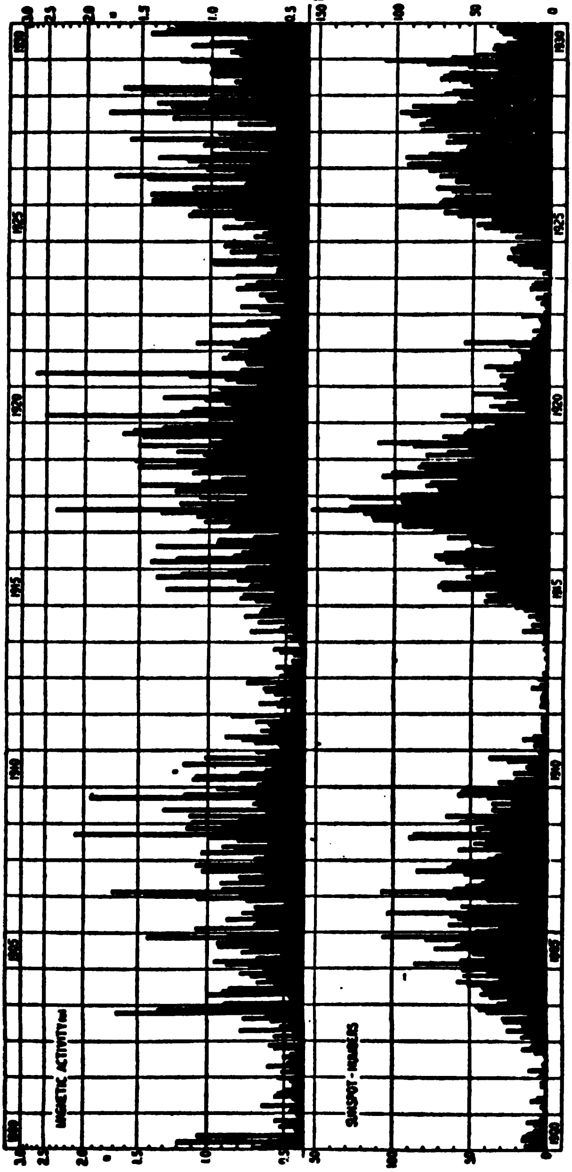


FIG. 29. Correlation between monthly means of magnetic activity (u) and sunspot numbers. (After Bartels.)

Besides this long period sunspot-cyclic variation of magnetic activity the incidence of magnetic disturbance shows an annual variation. Fig. 30 shows how the frequency of magnetically disturbed days varies throughout

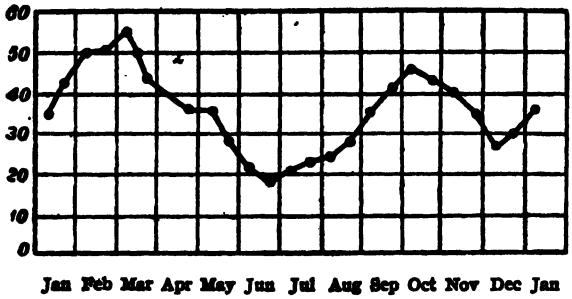


FIG. 30. Seasonal variation of frequency of magnetically disturbed days.

the year [18]. Two maxima are clearly seen one near the vernal equinox and the other near the autumnal equinox. Fig. 31 shows the variation of magnetic activity throughout the year for three different years in three epochs of solar activity. The upper one is the average for the most dis-

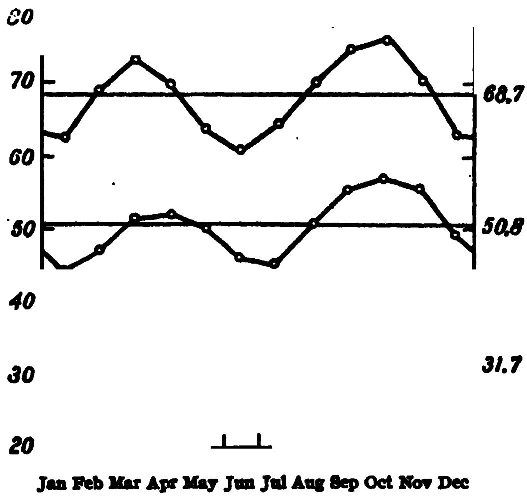
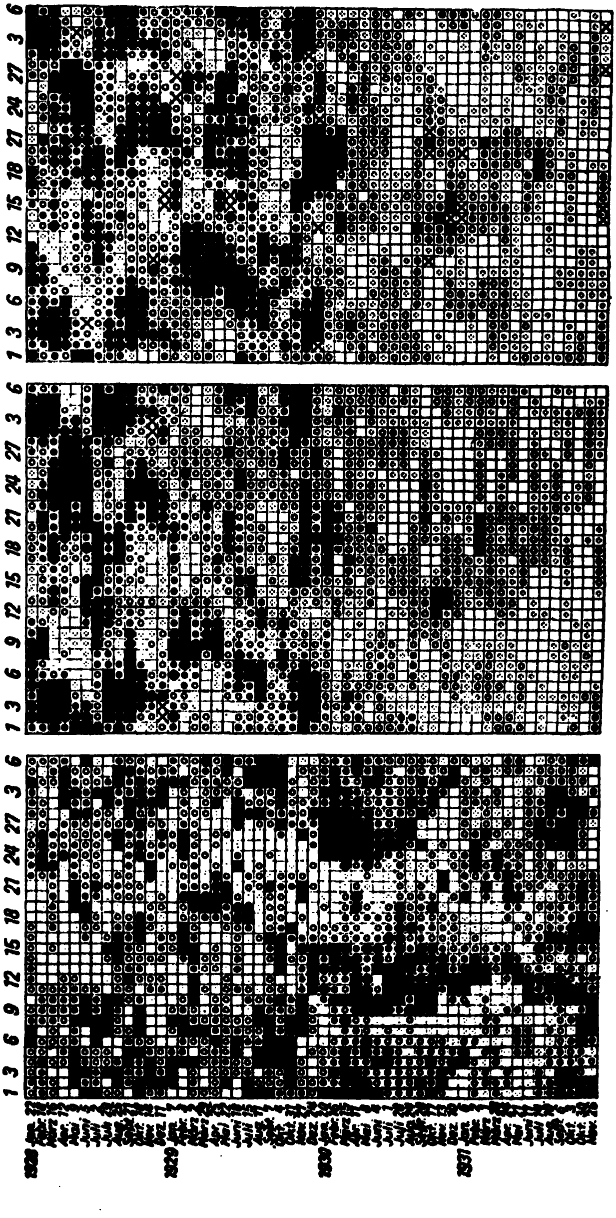


FIG. 31. Seasonal variation of magnetic activity (u) in most disturbed years (upper curve), years of average disturbance (middle curve) and quiet years (lower curve). (After Bartels.)

turbed years, the middle one for average years and the lowermost for quiet years. It will be noticed from the two figures that both the frequency and the intensity of the magnetic disturbance vary similarly throughout the year [18].

Besides the eleven-year cycle and the annual variation, the magnetic activity has another periodicity. The magnetic disturbance has a tendency to recur in about 27 days, i.e., after a solar rotation [52, 53]. This tendency cannot be described as a permanent periodicity because, not infrequently,

the recurrences may be interrupted and may sometimes cease altogether. The recurrence tendency may be depicted in various ways; as for instance, by means of the representation in Fig. 32(*a*). Different shadings of the square blocks in the figure depict different daily magnetic character figures. The most magnetically disturbed days are indicated by black squares and the quietest days by white blank squares. Intermediate shadings represent days of intermediate activity. The squares for the successive days are



(*a*) (*b*) (*c*)

0-□, 1-▣, 2-▣, 3-▣, 4-■, 5-■

FIG. 32. Illustrating the 27-day recurrence tendency of (*a*) daily magnetic character figure, (*b*) daily spottedness of the sun, and (*c*) intensity of hydrogen flocculi. (After Bartels.)

arranged in rows of 27. Thus, all the square blocks in one column refer to magnetic activity of days at intervals of 27 days. (For better comparison the magnetic activities of the first six days in any row are also given at the end of the previous row, i.e., after the 27th square.) It will be noticed that the black squares in successive rows tend to group themselves near the same column that is, at intervals of 27 days. The same remark applies to the blank or lightly shaded squares.

It may seem from the above that the mean daily magnetic activity should also show some correspondence with the daily spottedness of the sun. This, however, is not the case, as will be seen from Fig. 32(*b*) which depicts the daily spottedness of the sun in a manner similar to Fig. 32(*a*) for the daily magnetic character figures. Here again dark squares represent days of greatest spottedness and the white ones days of least spottedness. The arrangement in rows and columns is the same in both (*a*) and (*b*). It is easily seen that there is no similarity in the pattern of black, grey or white squares between Figs. 32(*a*) and (*b*), though, of course, in each of the figures the black or the white squares have a tendency to group themselves near the same column.

In view of the fact that bright hydrogen eruptions are found to be accompanied by radio fade-outs and also by terrestrial magnetic variations, it is interesting to compare the average number and intensity of hydrogen flocculi with the magnetic character figures [54]. In Fig. 32(*c*) the dark squares represent days with largest number of flocculi and the white ones days of least flocculi. It will be seen that while the patterns (*b*) and (*c*) show marked correspondence (as it should according to the well-known fact that flocculi and prominences are most numerous when the sunspots appear in large numbers) there is no such correspondence between the patterns (*a*) and (*c*). This fact is in accord with the observation that the sudden and intense changes in the terrestrial magnetic field, which are sometimes observed to accompany extra-brilliant hydrogen eruptions are not of the magnetic storm type. It has been found that such magnetic disturbances correspond to an intensification of the ionospheric current system which causes quiet day variations. The intensification is produced by penetrating ultraviolet radiations emanating from the eruptions. This has been discussed in Chapter VI, Sec. 13(*d*).

8. SOME MINOR VARIATIONS

(*a*) Magnetic bays

On magnetically quiet days, the magnetic elements occasionally undergo sudden variations. The values of the elements rapidly increase (or decrease) and, after attaining a maximum departure, regain their normal values in course of an hour or two. Since the magnetographic records of these variations resemble indentations on a sea coast, they are called 'bays' [55]. The bays often accompany magnetic storm variations—a great many occurring during a single storm; frequently one bay is partly or entirely

superimposed upon another. The characteristic features of the bays and their probable current systems are discussed below briefly [56].

The bays are called positive or negative according as the value of H increases or decreases.

The characteristics of the bays depend markedly on the latitude of the place of observation. Near the centre of the auroral zone disturbance in the Z component is relatively much smaller than in the H component. Near the auroral zone both H and Z variations are great, the amplitudes depending in a marked way on the time of the day. In middle latitudes the east component rather than the north component is affected, while in the equatorial regions the north component is the most marked one.

The bays are generally preceded and followed by undisturbed magnetic conditions. Though the bays are most pronounced near the auroral zones, their magnetic effects extend to equatorial regions.

The maximum departure for H is of the order of 5 to 20γ in middle latitudes. The departure of Z is much less—about 1 to 3γ. The declination is less affected. In higher latitudes the departures for H and Z might be 10 times as great.

The positive bays are more frequent than the negative bays, the ratio being 4 : 1 to 2 : 1. The former occur mostly during night (during the eight hours centred at midnight), and the latter mostly during day in the afternoon [57, 58, 59].

A large bay occurs simultaneously all over the world; its form and magnitude depend on its geographic position and time.

An intensive study of the bays was made during the Second International Polar year 1932-33 with collaboration of many countries [60]. This has led to a clearer understanding of the nature and immediate cause of the bays. The observed features of the bays may be attributed to the sudden flow of a current in the upper atmosphere, along a long narrow path parallel to the auroral zone. In the night side of the zone (northern) the current sheet flows westward. In some cases it completes the circuit by turning southward at the sunset meridian, eastward at the tropics and northward at the sunrise meridian. The field due to such a current will obviously produce a negative bay near the auroral zone and a positive bay in the middle latitudes. Such association of negative bays in high latitude with positive ones in the low latitudes is known to exist. It now seems that the intense and frequent negative bays in high latitudes are to be identified with the type of magnetic disturbance first noticed by Birkeland [61] and called by him *elementary polar* storms. Chree's and Birkeland's [55, 61] observations, namely, bays in middle latitudes are often accompaniments of elementary polar storms in high latitudes, have since been confirmed by other workers.

A quantitative estimate of the upper atmospheric current system which could have produced a bay disturbance as observed on January 15, 1933, at 21 h. 45 m. G.M.T. at a number of stations near the auroral zone, has been made by McNish [60, 62]. In Fig. 33, the observed variations are plotted against latitude for both X and V. Idealized arbitrary variation curves

are also drawn through the observed points. The intensity distribution across a current sheet, situated at a height of 100 km., which could produce the idealized variation curves, is shown in Fig. 34. The height of the current sheet was also estimated after the method indicated in Sec. 6(*f*).

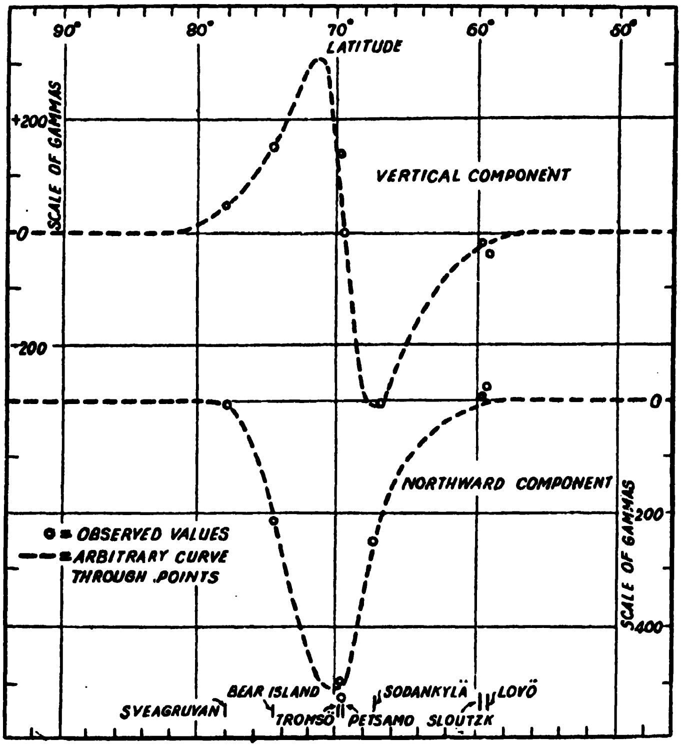


FIG. 33. Variations with latitude of X and V during a bay disturbance observed on January 15, 1933, at 21 h. 45 m. G.M.T. (After McNish.)

There is, however, some uncertainty about the result as the effect of the induced earth current could only be allowed for approximately. Making the probable allowance the height was estimated to be 66 km.

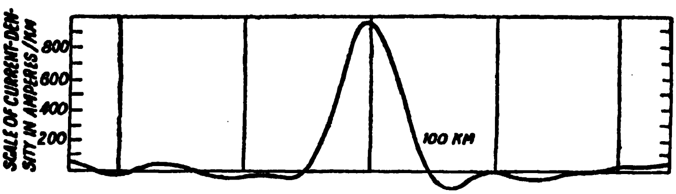


FIG. 34. Intensity distribution with latitude of westward current sheet at a height of 100 km. which could have produced the magnetic bay at 21 h. 45 m. G.M.T. on January 15, 1933. (After McNish.)

An interesting point regarding the height of the 'bay' current system has been indicated by Wells [63]. According to Wells' observations

there is direct relationship between occurrence of the bays and radio 'black out' [see Chapter VI, Sec. 13(*d*)]. Since, for producing the marked absorption of radio waves leading to a black out, the increase of ionization should occur below 80 km., the current system producing the magnetic bays must also be concentrated at such levels.

(*b*) Micro-pulsations

Another interesting type of variation of magnetic elements is the so-called micro-pulsations. Magnetic records of all the three elements sometimes show rapid fluctuations in the form of oscillations lasting for an hour or so. The period varies between a fraction of a minute and 3 minutes and the amplitude is several gammas. In extreme cases the amplitude may be as great as 15 to 30γ. Pulsations of such large amplitude are called giant micro-pulsations.

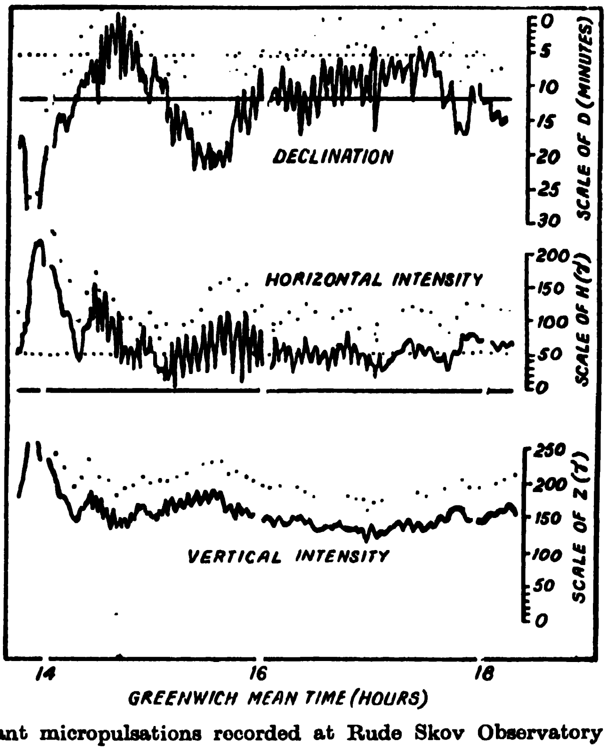


FIG. 35. Giant micropulsations recorded at Rude Skov Observatory on March 1, 1942. The discontinuities in the records are produced by displacements of the photographic paper to mark even hours.

The characteristic features of a typical giant micro-pulsation as recorded at Rude Skov Magnetic Observatory on March 1, 1942, at 14–18 hour G.M.T. are shown in Fig. 35 [64]. Harang [65] has closely studied another giant micro-pulsation reported earlier by Rolf [66]. An interesting feature observed about these pulsations is that the oscillations of the different

elements are not in phase. The oscillations of the three elements may be written in the form:

$$\delta W = A_d \cos 2\pi \frac{t}{T},$$

$$\delta H = A_h \cos \left(2\pi \frac{t}{T}+\alpha\right),$$

$$\delta V = A_v \cos \left(2\pi \frac{t}{T}+\beta\right),$$

so that the variation of the total force P is given by

$$\delta P = \sqrt{(\delta W)^2+(\delta H)^2+(\delta V)^2}.$$

The following are the values of the constants for *one* of the observed pulsations:

$$A_d = 6{\cdot}5\gamma,\ A_h = 5{\cdot}7\gamma,\ A_v = 4{\cdot}2\gamma,$$

$$\alpha = -90^\circ,\ \beta = -120^\circ,$$

$$\text{Period } T = 115 \text{ sec.}$$

The giant pulsations occur within a relatively limited area and their periods generally range from 1 to 2 minutes. The local character of such pulsations indicates that they must be due to some peculiar type of currents flowing in a limited region of the upper atmosphere above the place where the pulsations are observed with greatest intensity.

Besides such giant pulsations, others of much less intensity and very short periods (5 to 15 sec.) also occur at different parts of the earth. These small micro-pulsations were first observed by Eschenhagen [67] and later extensively studied by Angenheister [68]. They seem to occur simultaneously all over the world. For such micro-pulsations the vertical intensity is little affected—at least in lower and middle latitudes. In higher latitudes the vertical intensity is affected but the amplitude is much smaller than that of the horizontal components. The pulsations usually occur around midnight and show a tendency to recur at an interval of 27 days.

A possible origin of these pulsations has been suggested by Störmer in course of his extensive studies of trajectories of charged particles in the magnetic field of the earth. Störmer discovered that under certain conditions the particles (which are supposed to be ejected from the sun and to be the ultimate cause of terrestrial magnetic disturbance) may move in periodic orbits far out in space (See Chapter IX, Fig. 13). Further, for certain types of orbits the calculated periods are of the same order as those of the rapid micropulsations. Störmer therefore suggests that such pulsations are manifestations of the effect on the terrestrial magnetic field, of charged particles moving in these periodic orbits.

CHAPTER VIII

AURORA POLARIS

1. INTRODUCTION

The striking luminous displays observed in the high atmosphere of the polar regions known as *aurora polaris* are also called *aurora borealis* in northern countries and *aurora australis* in the southern countries. The aurora borealis or northern light has been studied more extensively and over longer periods, because of the easier accessibility of the northern polar regions than the southern ones. There is little doubt that if observations are carried out in southern regions the aurora australis will also be found to possess characteristics similar to those of the aurora borealis.

The auroral displays have a great variety of forms. According to Störmer they can conveniently be grouped into two main classes: (1) Those exhibiting a ray structure, and (2) those without such structure. The former class includes the *corona*, the *rays* and the so-called *draperies*, while the latter comprises *homogeneous arcs*, *homogeneous bands* and *pulsating surfaces*. The classification may also be made on other basis. Vegard, for instance, distinguishes auroras as having quiet forms and of moving types. The draperies, the rays, the ray-bundles and the coronas are generally of the moving type. Brief descriptions of the various auroral types are given below. Each type is designated by a symbol according to the classification adopted by the International Geodetic and Geophysical Union.

The *arcs* may be with or without ray structure. In the latter case they are called homogeneous arcs (symbol *HA*). They have fairly sharp boundaries and extend across the sky in some direction which is typical of the locality. The highest point of an arc is on the magnetic meridian. With bright arcs the upper part is green, the middle yellow and the lower part usually red. The arc is often accompanied by rays which appear to diverge from it like spokes of a fan (symbol *RA*). The arcs may also sometimes be pulsating, flashing up and disappearing in a period of a few seconds (symbol *PA*).

The *rays* (symbol *R*) may appear singly or in bundles in great masses. Sometimes they are quiet, without any movement, only appearing and disappearing; at other times they may be in rapid motion.

The *draperies* (symbol *D*) have a curtain-like appearance with very long rays. Sometimes the rays of the drapery follow the magnetic lines of force and have a fan-like appearance.

The *crown* or the *corona* (symbol *C*) is seen further north where the lines of force are nearly perpendicular to the surface of the earth. The ray-streamers of the corona are seen to spread out from the particular point

in the sky which lies near the magnetic zenith of the observer. [See Sec. 2(*d*).]

The *bands* may be homogeneous (symbol *HB*) or, with ray structure (symbol *RB*). They extend in the same direction as the arcs.

The *diffuse luminous surfaces* (symbol *DS*) have the appearance of luminous clouds of indefinite shape and indistinct boundaries. They may also be sometimes pulsating with the intensity undergoing fluctuations within periods of about 8 to 10 seconds (symbol *PS*); they appear as if diffuse clouds are being illuminated by a source of light the intensity of which is rapidly changing.

The upper part of the arc may sometimes appear as a feeble *glow* (symbol *G*) near the horizon resembling dawn.

Of these various forms the most common are the arcs with or without ray structure, the bands and the draperies; pulsating surfaces and rays are comparatively rare. Photographs of some typical auroral forms are reproduced in Plate I. These are selected from the *Photographic Atlas of Auroral Forms* published by the International Geodetic and Geophysical Union, Oslo (1930). (*a*) Homogeneous quiet arc. Note that there is a dark segment between the arc and the horizon; Konigsberg, 16-17 October, 1923. (*b*) Masses of rays; Bygdö, 5-6 March, 1926. (*c*) Draperies; Bygdö, 15-16 October, 1926. (*d*) Corona. Only half has been developed; Bygdö, 16-17 December, 1917. (*e*) Band with ray structure; Bossekop, 3-4 March, 1910. (*f*) Homogeneous band with many folds; Korsnes, 15-16 March, 1913.

The study of auroras is of great importance in the investigation of the upper atmosphere. Spectroscopic observations of the auroral light in particular, yield valuable information regarding the composition and temperature of the upper atmosphere. The study of the form and the geographical distribution of the auroras is again very helpful in investigating the nature and origin of the charged particles, the entry of which into the high atmosphere is the cause of the magnetic storms and also of the auroral displays. It is satisfactory to note that to the visual, photographic and spectroscopic methods of studying the auroras, has now been added the new and powerful radar method. Radar echoes from auroras, due to the intense ionization produced by the bombarding solar corpuscles, have been recorded by more than one observer [1, 2].

2. CHARACTERISTICS OF AURORAS

(*a*) Geographical distribution

As already mentioned, auroras occur mainly in the regions around the magnetic poles of the earth; the frequency of their displays increases as one proceeds from lower to higher magnetic latitude. Fig. 1, prepared by Fritz [3] from numerous observations made in the arctic regions, shows the variation of auroral frequency with both latitude and longi-

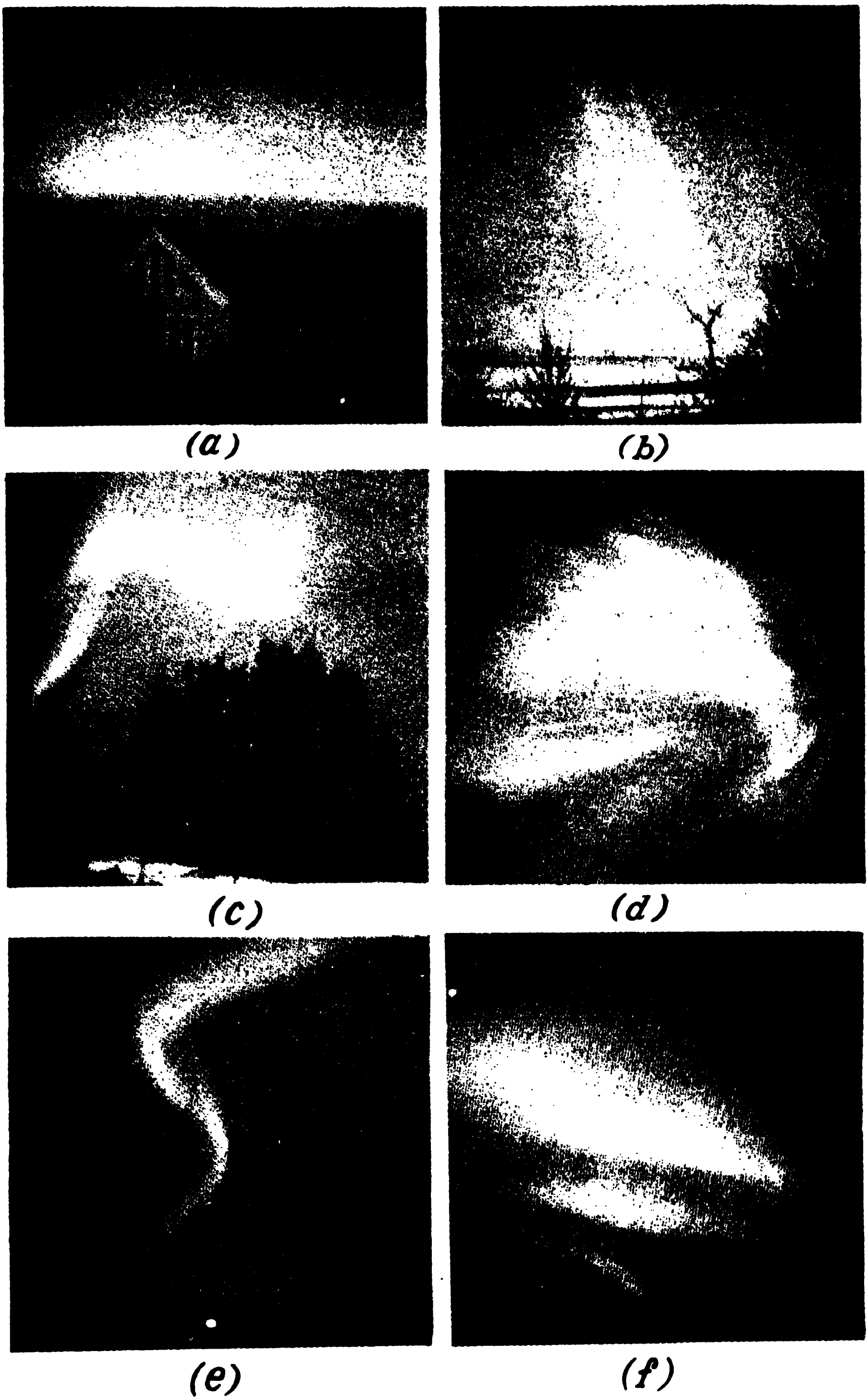


PLATE I. Photographs of some typical auroral forms. (See text.)

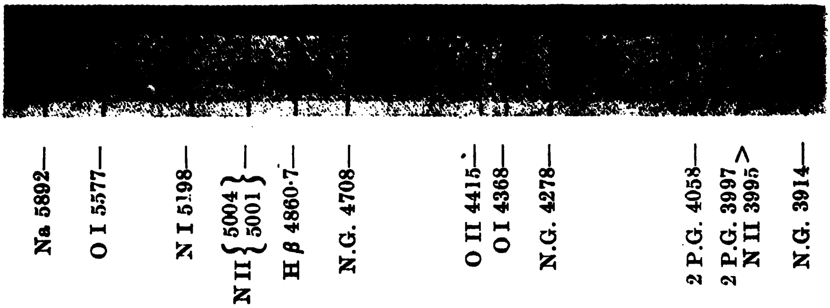


FIG. 15. A typical auroral spectrum. (After Vegard.)

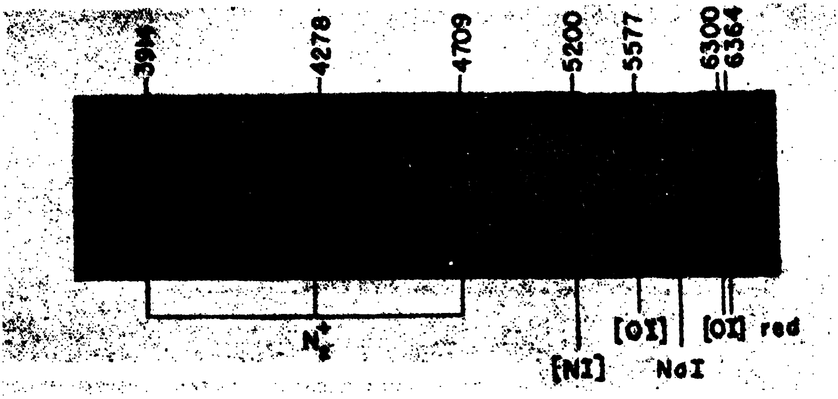


FIG. 17. High altitude auroral spectrum. (After Dufay and Tcheng Mao-Lin.)

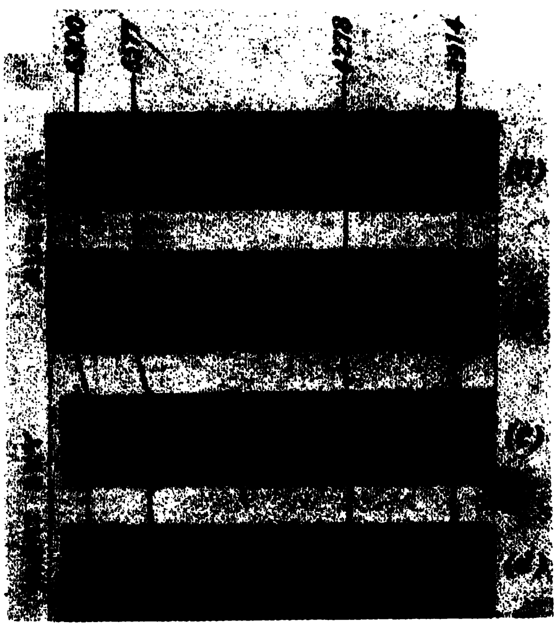


FIG. 20. Illustrating the enhancement of the red oxygen lines and of the 1st negative bands of N_2^+ in sunlit auroras and in the twilight spectra of the upper atmosphere. The night sky spectra were taken by Elvey (1942) at the McDonald Observatory, Texas, U.S.A.

The auroral spectra were taken by Störmer on September 15, 1938, in Southern Norway.

tude. The lines of equal frequency or, *isochasms* as they are called are seen to have roughly circular form with their centres lying approximately at the point where the axis of the dipole representing the regular part of Earth's magnetic field cuts the surface of the Earth. The thickest line in the figure corresponds to the zone of maximum frequency which is approximately in the region 67° N. magnetic latitude. As one travels towards north from this region the auroral frequency decreases. In the region indicated by the broken line auroras are seen in all directions. In all other lower regions the displays occur to the north of the observer. It is to be mentioned that the diagram of Fritz has been revised and improved by Vestine on the basis of much subsequent data [4].

A zone of maximum frequency, similar to the one discussed above, may be expected to exist in the southern polar region. From the meagre observational data which are available White and Geddes [5] have attempted to draw such a zone (Fig. 2). The zone is incomplete but there is little

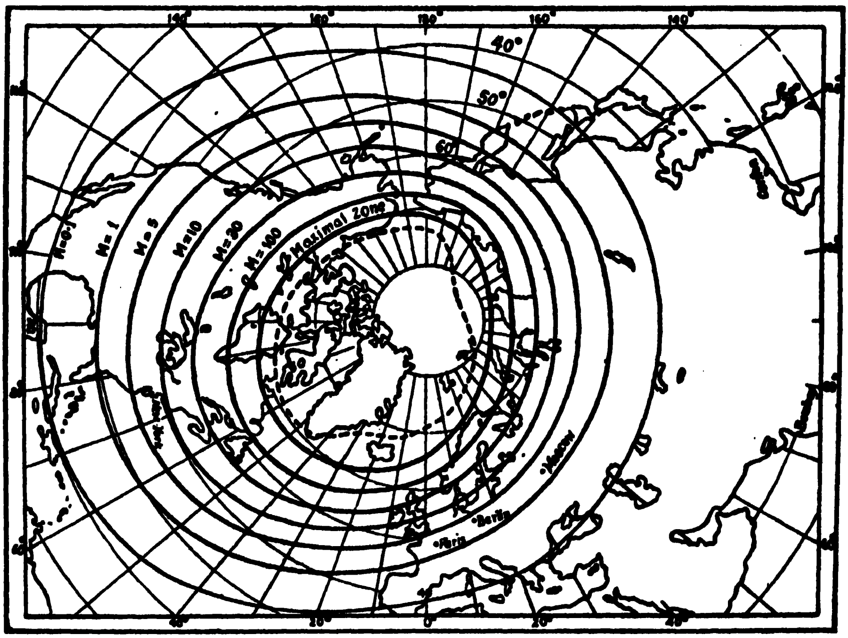


FIG. 1. Illustrating after Fritz the variation of the frequency of auroral displays with latitude and longitude in the Northern Hemisphere. The regions of equal frequency are shown by the thick lines, the thicknesses being proportional to the frequency. The numerical figures after *M* give the average relative frequency of auroral display. It is to be noted that lines of equal frequency are roughly of circular form and their centres lie close to the magnetic axis point. In the region indicated by the broken line auroras are seen in all directions; in all lower regions the display occurs to the north of the observer.

doubt that if more data were available it would have formed a complete circle. The zone is very nearly circular with a radius of about 18°. It

is satisfactory to note that its centre lies very close to the antipodes of the northern magnetic axis pole [6].

(b) Altitude distribution

Fig. 3 represents the distribution of auroras with altitude computed from a large number of observations made by Störmer [7] from his various observational stations in Norway and also by Vegard and Krogness [8] at Haldde observatory. It will be noticed that nearly all (about 94 per cent) the auroras occur between 90 and 130 km. The dotted curve due to Vegard and Krogness shows two maxima, one at 100 km. and the other at 106 km.

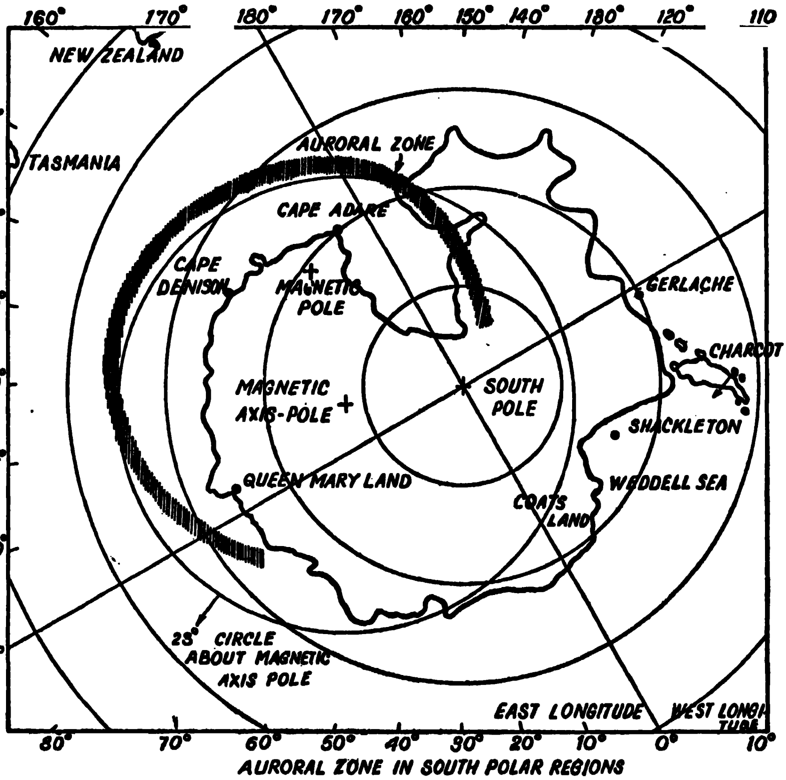


FIG. 2. Zone of maximum auroral frequency (shaded region) in the Southern Hemisphere. The centre of the zone lies close to the antipodes of the north axis pole. (After White and Geddes.)

According to Störmer the lower limit of auroral height is about 80 km. Though this is generally true, auroras at heights as low as 63 km. [9] are sometimes observed. Some reporters [10] claim to have observed auroras very close to the surface of the earth; close scrutiny reveals, however, that all such reports are groundless.

As regards the upper limit, ordinary auroras have been observed to occur up to a height of 400 km. Since the upper edge of most auroral forms is ill-defined the percentage of error in the determination of the

upper limit is rather high. Sunlit auroras, to which reference will be made presently, have generally much higher altitudes. The highest altitude as yet measured is about 1100 km. [11].

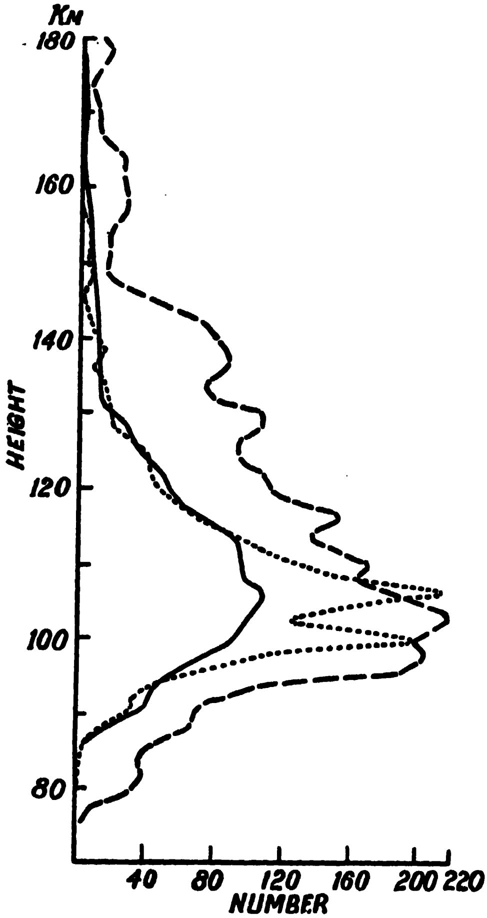


FIG. 3. Distribution with height of the lower limits of auroras as observed at three Norwegian stations. The dotted curve is computed from observations made by Vegard and Krogness at Haldde; the dashed curve from observations by Störmer at Oslo; the full line curve from observations of Störmer at Bossekop.

Tables I and II give the lower and the upper limits of auroral displays of various forms as measured by different observers [12]. Fig. 3 depicting the distribution of auroral heights is drawn from Table I. The two tables also show the relative frequency of occurrence of the various forms. It may be noticed that the most frequent forms are the arcs and the draperies. Table I further shows that there is no marked difference between the heights of different types of auroras, though it appears that the arcs and the pulsating surfaces have, on the average, slightly lower heights than draperies, and particularly than the rays. A comparison of the average heights with reference to the latitudes of the observational stations shows that auroras occur at lower heights in the lower latitudes. This is in conformity with the theory of auroras [see Chapter IX, Secs. 3, 4].

Measurement of height of auroras.—The height of an aurora was first accurately measured by Störmer [13] at Bossekop in 1910. He took two simultaneous photographs of the aurora from two stations separated by a distance of four and a half kilometres. The exact hour of exposure was known and the photograph also recorded a number of known stars in the background. The position of the aurora (i.e., of any particular point in it) in space was obtained by proper orientation of and computation from the negatives of such simultaneous photographs. The method has subsequently been improved with certain modifications by Vegard and Krogness [8]

TABLE I

Lower limits of auroral displays as measured by Norwegian observers: I—Vegard and Krogness at Haldde; II—Störmer at Bossekop; III—Harang and Tönsberg at Trömso; IV—Störmer at Oslo.

Auroral type	I		II		III		IV	
	Height (km.)	Number of observa-tions	Height (km.)	Number of observa-tions	Height (km.)	Number of observa-tions	Height (km.)	Number of observa-tions
Rays ..	113·2	61	128·9	19	117·0	127	146·9	119
Draperies ..	109·8	409			112·9	1039		
Drapery-shaped arcs	106·6	888	108·1	506	106·7	1175	100·0	150
Diffuse arcs	109·1	409	111·8	415			118·5	201
Pulsating surfaces ..	106·0	160			107·3	66		

TABLE II

Upper limits of auroras observed at the Haldde Observatory [12].

Auroral type	Number of observations	Height (km.)	Average vertical extension (length of streamers) (km.)
Rays	43	250	137
Draperies ..	215	176·3	67
Arcs with ray structure	174	174·4	68
Homogeneous arcs ..	57	143·4	34

and by Harang and Tönsberg [14]. Herlofson [15] has devised means for simplifying the tedious calculations involved in the usual method of height computation. Chapman [16] has proposed a very simple and interesting method for analysing the simultaneous photographs by exactly reproducing the original states of cameras and the aurora by projecting the negatives, with two projectors, on a rigid framework representing the celestial sphere. The positions of the known stars at the particular time of exposure are marked on the frame work and the projectors are oriented, in turn, to such positions that stars projected from the negatives coincide with those on the frame work. The positions of the two projectors represent the

two stations and hence the exact height of any point on the aurora can be found by actual measurement and calculation.

At present regular height measurements are made from a network of auroral stations in southern Norway under the direction of Störmer. Simultaneous photographs are taken from three or four stations. The base line joining the various stations vary from 26 to 400 km. Besides the regular observations by Störmer, other notable sets of measurements are by Vegard and Krogness [17] at the Haldde observatory and by Harang and Tönsberg [18] at Tromsö. Some measurements have also been carried out in northern Canada by McLennan, Wynne-Edwards and Ireton (lat. 50° 40′ N., long. 81° 25′ W.) and by the Canadian Polar Year groups at Chesterfield [19]. The average heights as measured by the above observers all refer to the ordinary auroras. The so-called sunlit auroras to which reference will be made presently attain great heights, the base being at 200 to 300 km. and the tops vanishing at 800 to 1000 km.

Since, as already mentioned, many auroral forms have a sharp bottom edge the determination of the height of the lower limit of the aurora is much easier and can be made with an accuracy of 1 to 2 per cent when the zenith distance is not much greater than 45°. For greater zenith distances the error is comparatively greater and may be as high as 20 per cent.

(*c*) Azimuths of arcs, bands and draperies

Arcs, bands and draperies tend to extend in a horizontal direction roughly *perpendicular* to the magnetic meridian (Fig. 4) in contrast to the streamers which lie more or less *along* the lines of force. The actual direction varies with the locality and is somewhat characteristic of the place of observation. In Fig. 4, the horizontal projections of some homogeneous arcs as photographed from Bossekop and Oslo are drawn. The broken lines are circles of equal distance from magnetic axis pole (78·5° N., 69° W.). It will be noticed that the projections do not strictly coincide with the magnetic latitudes. According to measurements of Vegard at Bossekop there is a systematic difference of 10°. The westward ends of the arcs are thus somewhat nearer to the pole than the eastern ends. This result has important bearing on the question regarding the sign of the electric charge carried by the solar corpuscular rays to which reference will be made later [see Chapter IX, Sec. 4(*b*)].

It was also noted that the above rule holds even for observational stations situated near the ordinary magnetic dip-pole. At stations situated close to the magnetic axis point, however, the direction of the bands was found to be irregular. The obvious interpretation of this is that the distribution of the direction of arcs and bands is to be referred to the magnetic axis pole rather than to the magnetic dip-pole.

The arcs often extend to great lengths. The arc for instance shown in Fig. 5 was observed on March 24, 1936, to extend from longitude 17° E. to 3° W. lying between latitudes 63·5° and 65·5°, partly over Norway and partly over the North Sea. Its projection on the surface of the earth is

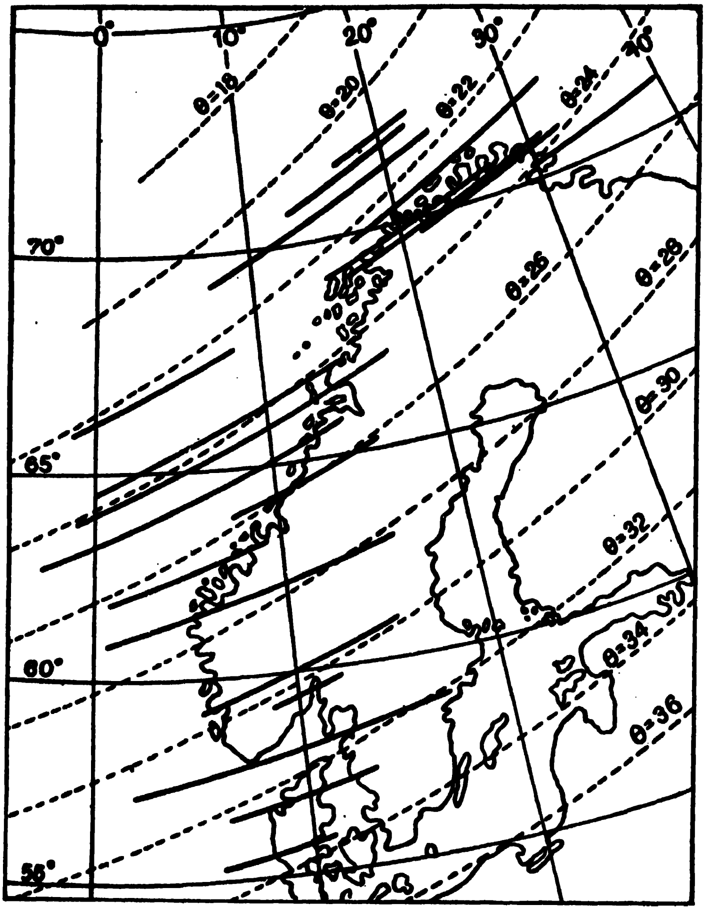


FIG. 4. The horizontal projections of arcs nearly coincide with magnetic latitudes. The broken line curves are circles of equal distance from the magnetic axis pole assumed at 78·5° N., 69°·W. The heavy full lines are the projections. The systematic deviation of the directions of the arcs has important bearing on the question of the sign of charged particles producing the aurora. (After Störmer.)

shown in the figure. (This is the most probable size and position as obtained from three assumed heights namely, 105 km., 95 km. and 85 km. Owing to unfavourable weather conditions observation from more than one station—which is necessary for the computation of the height—could not be made [20].)

Note.—For auroral and other magnetic phenomena it is necessary to refer the observations to the magnetic axis point. The points may be referred to the magnetic co-ordinates—magnetic longitude, latitude and azimuth. The magnetic longitude is measured from the great circle passing through the geographical north pole and the north axis point. One can easily pass from astronomical to magnetic co-ordinates and *vice versa*, when the co-ordinates (ϕ_0, λ_0) of the magnetic axis point are known.

For instance, if α is the astronomical azimuth and a_m the magnetic azimuth then

$$a_m = \alpha - A_m,$$

where A_m is the azimuth of the magnetic meridian referred to the axis point.

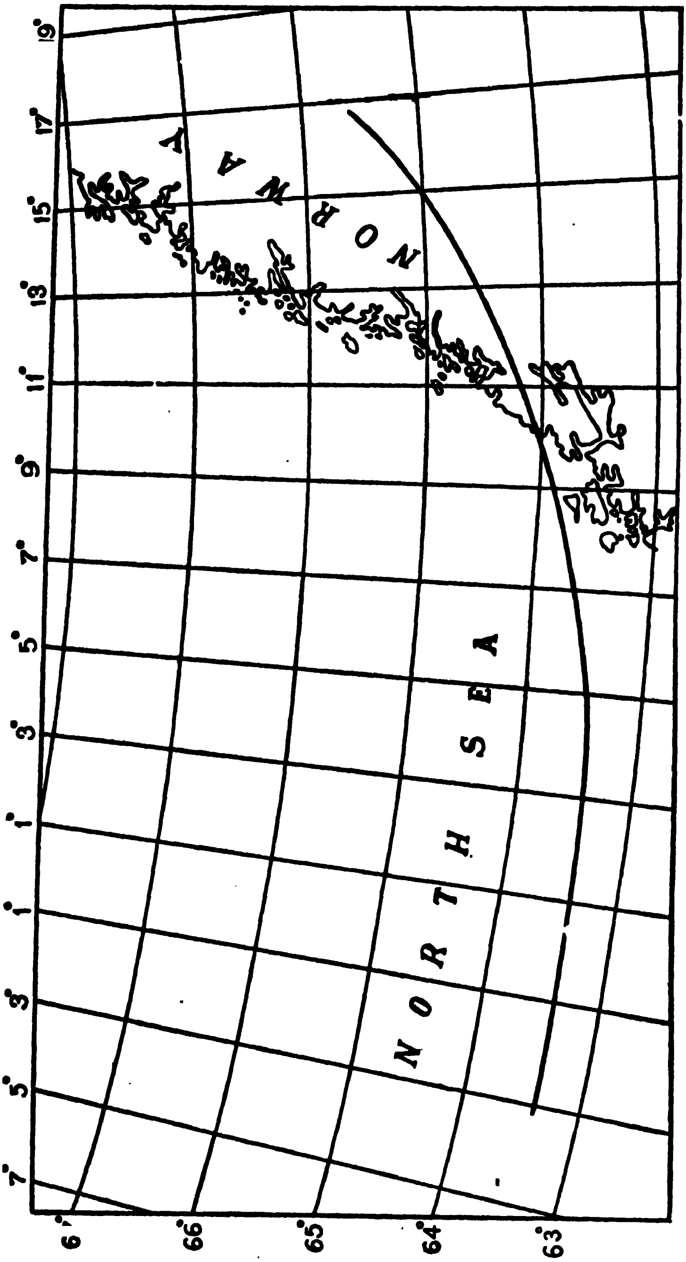


FIG. 5. Illustrating the great length of the arcs. The arc in the illustration was observed on March 24, 1936 (After Störmer.)

(*d*) Position of the radiation-point of corona

It has already been mentioned that the directions of the auroral streamers approximately coincide with the lines of force of the permanent

magnetic field of the earth. As a matter of fact, the radiation-point nearly coincides with the *magnetic zenith*, i.e., the point where the direction of the lines of magnetic force at the place of observation, meets the celestial sphere. Careful measurements show however that the radiation-point is situated a few degrees south of the magnetic zenith [8]. Now, a study of the distribution of luminosity and the structure of the auroral streamers suggests that the ray streamers ought to coincide with the direction of the lines of force of the magnetic field. In other words, the direction of the streamers and the position of the radiation point, as observed during a display should mark respectively the direction of the magnetic lines of force and the instantaneous position of the magnetic zenith on the celestial sphere at the time and the place of observation. And, since the observed radiation point is situated a few degrees south of the magnetic zenith, it may be concluded that the auroral display has, for the time being, so changed the direction of the lines of force that the magnetic zenith has been shifted to coincide with the observed radiation-point. This change in the magnetic field and the consequent shift of the magnetic zenith may be due to extra-terrestrial current systems generated during auroral displays. We shall see later, while discussing the theory of magnetic storms [Chapter IX] that the extra-terrestrial current systems which may be responsible for producing the storms, are such as to cause an increase in the curvature of the magnetic lines of force and produce a lowering of the magnetic zenith.

(e) Intensity distribution with height

The intensity of light along the auroral streamers varies greatly from point to point. The following method of characterizing the intensity distribution is due to Vegard [21]. Three quantities l_1, l_2 and l_3 are associated with the streamers (Fig. 6); l_1 is the distance of the point of

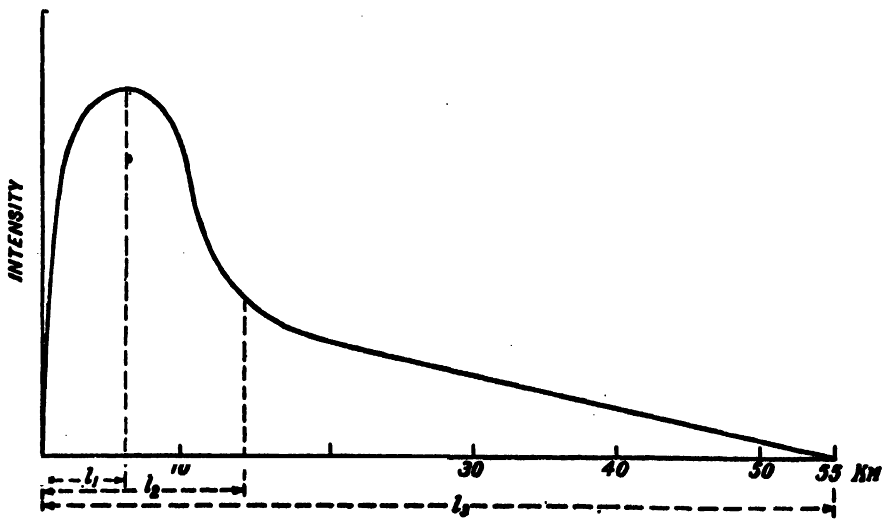


FIG. 6. Characterising the intensity distribution in auroras after Vegard.

maximum intensity from the bottom edge, l_2 is the distance from the same of the point where the intensity just becomes faint and l_3 is the whole length of the luminescence. The lengths l_1 and l_2 can be determined with fair accuracy; l_3 however is more difficult to estimate.

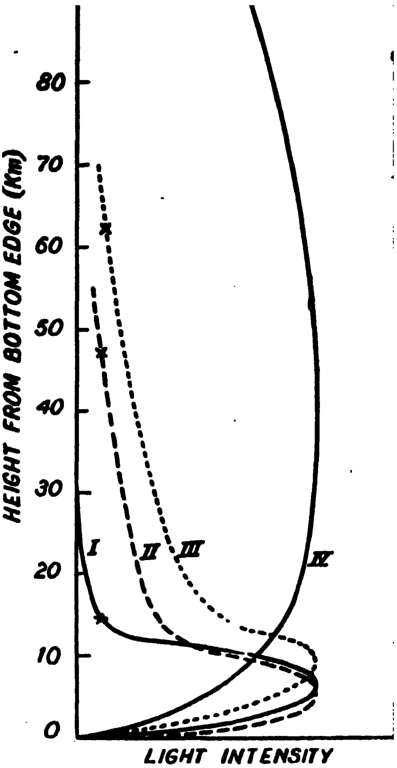


FIG. 7. Variation of intensity with height of different auroral types. I—Arc; II—Drapery-shaped arc; III—Drapery; IV—Rays. The crosses indicate upper limits of photographic impression. (After Vegard and Krogness.)

Fig. 7 shows the vertical distribution of intensity for some typical auroral forms. It will be noticed that for arcs and drapery-shaped arcs l_1 and l_2 are very small. In fact, the range of luminescence is restricted within an interval of about 14 km. only. For the arcs l_2 and l_3 can hardly be distinguished, i.e., the luminosity drops down quite suddenly. For the rays l_2 is very long, i.e., the entire length is faintly luminous with fairly constant intensity. The very small cross-section which the ray streamers sometimes have is a point of special interest. Measurements [21] show that in some cases the width of the cross-section may be only 300 to 400 metres.

(*f*) Diurnal variation of auroral activity

For the purpose of studying the variation of auroral activity, it is necessary to provide a quantitative measure for it. One of the methods of measuring the auroral activity is to divide the period of observation into half hour intervals and count the number of displays in each interval. This method has the disadvantage that weak as well as strong displays get the same weight. Also, near the auroral zone the activity, so measured, may show a uniform maximum throughout the whole period of observation.

A better and objective method of studying the variation is to register the total luminosity by means of a photoelectric cell. This method has been employed by Harang and Kreielsheimer at the auroral observatory at Tromsö, and also by Dauvillier. A spectrographic method, due to Harang [22], is to measure the average intensity of the green auroral line at intervals of one hour. The method has the advantage that it allows recording of auroral intensity even during the hours of moderate daylight.

Careful analysis of observations made from various stations by different workers shows that the auroral activity has two diurnal maxima. This is illustrated in Fig. 8 after Currie and Edwards [23]. One of these, called the principal maximum, is very prominent and occurs before midnight, and the other maximum, a weak one—occurs in the early morning. The principal maximum, in particular, has been very carefully studied by Vegard [24] by analyzing the records of the draperies, rays and coronas as obtained from a number of stations during the Polar Year 1882-83. He found that

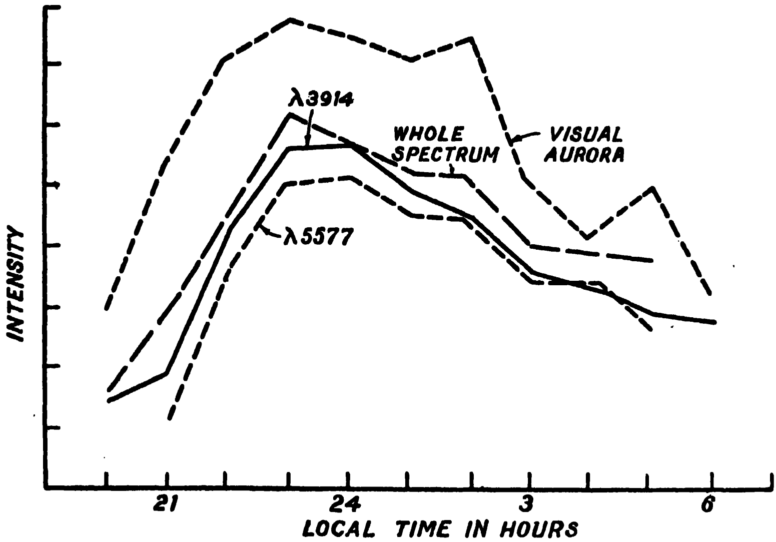


FIG. 8. Diurnal variation of auroral intensity as observed by Currie and Edwards at Chesterfield, Canada (63° 20′ N., 90° 42′ W.).

the hour of the principal maximum is distributed comparatively widely if referred to the astronomical time but very much narrowly, if referred to the magnetic local time, i.e., referred to the magnetic axis point as the pole. From this Vegard concludes that 'within the limit of error the principal maximum occurs at the same magnetic local time, about one hour before magnetic midnight'.

The midnight maximum is caused mostly by strong moving auroral forms whereas the weak early morning maximum is produced by quiet forms.

An interesting point in connection with the variation of intensity, is the variation of the geographical position of the aurora. According to Vegard [24] there is an intimate correlation between the intensity and the position of the place of observation. For instance, for a place south of the auroral zone (in the northern hemisphere) auroras appear in the north in the early evening and draw towards the south as the night advances. During the hours of highest intensity, i.e., near about the magnetic midnight the auroras appear in the most southern position. The auroras draw northward again after midnight. This is illustrated in

Fig. 9 after Carlheim-Gyllenskiöld from observations made at Cap Thordsen during the 1882-83 Polar Year [25]. It will be noticed that the relative

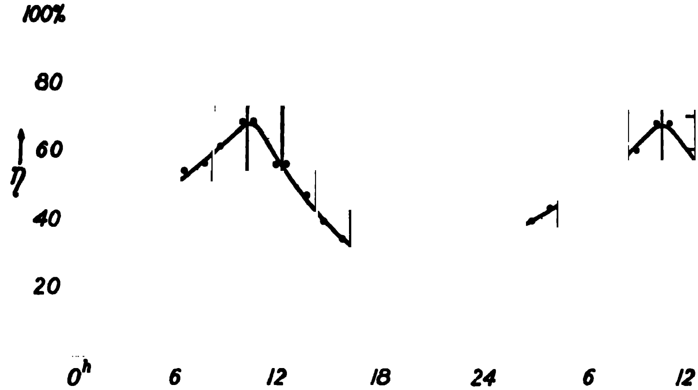


FIG. 9. Relative numbers of auroras in the northern sky. η is the ratio of the auroras in the zenith (or to the north of it) and the total number of auroras. Observations made by C. Gyllenskiöld at Cap Thordsen.

number of auroras in the northern sky is highest in the morning hours and lowest round midnight when most of the auroras occur in the southern sky. (In the figure the abscissae indicate local time. If the magnetic time is used better agreement is obtained.) The phenomenon has thus been summarized by Vegard: An increase of intensity is accompanied with a motion towards the lower latitude. We shall see later that this is compatible with the theory of aurora, namely, that auroral displays are caused by entry of fast charged particle in the upper regions of the atmosphere and that their motion is controlled by the magnetic field of the earth.

(g) Seasonal variation of auroral activity

The auroral frequency has also an annual periodicity as illustrated in Fig. 10. Like the magnetic activity, it has, in lower latitudes, two maxima

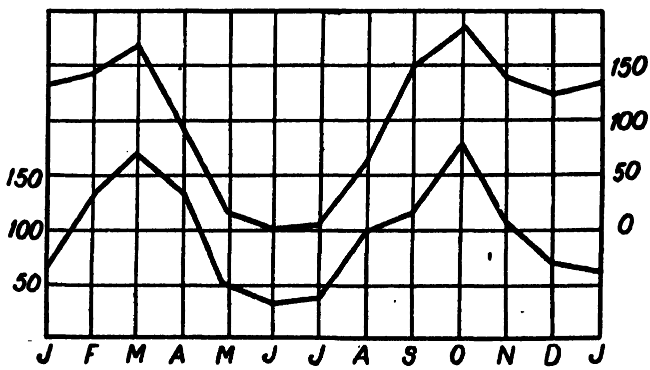


FIG. 10. Annual variation of the auroral frequency. The upper curve is for Sweden and the lower curve for Southern Hemisphere. Annual mean ordinate is taken to be 100 for each curve.

coinciding with the equinoxes. Unlike the magnetic activity, however, the two maxima approach each other as one proceeds towards the auroral zone where there is only one maximum occurring in mid-winter. Attempt is sometimes made to correlate the two maxima with the position of the earth relative to the two zones on the sun where the sunspots are most numerous. But, as remarked in Chapter IX, Sec. 2(*e*), closer scrutiny fails to reveal any correlation.

(*h*) Correlation between auroral activity, solar activity and magnetic storms

It has been shown by various observers that the auroral activity follows the well-known 11-year solar activity. This parallelism is illustrated after Boller [26] in Fig. 11. The variation is less pronounced near the auroral zone and becomes marked in lower latitudes. It has also been observed that auroras have a tendency to occur at the time when large sunspot groups pass near the central meridian of the sun [27, 28].

A remarkable similarity between the occurrences of magnetic storms and auroras is the 27-day recurrence tendency [29]. Strong auroral displays are not infrequently followed by similar displays at 27 days interval—the period of synodic rotation [see Chap. VII, Sec. 7; also Appendix, Sec. 8].

It is well known that earth current records are similar to the magnetic activity records. It has therefore been sought to establish correlation between auroral activity and magnetic activity by comparing the former with earth current records. This is illustrated in Fig. 12 after Currie and Edwards [30]. Observations [31] made at College Fairbanks Station at Alaska (lat. 64° 41′ N., long. 148° 25′ W.) show that the coefficient of linear correlation between the two phenomena may be as high as 0·71 to 0·76.

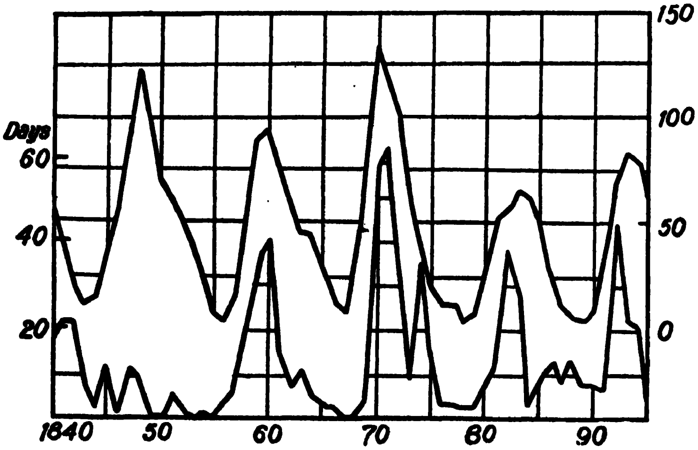


FIG. 11. Illustrating the parallelism between the auroral and the solar activity during the years 1840–95. The upper curve gives the relative sunspot numbers and the lower one, the numbers of days on which aurora australis was seen. (After Boller.)

In fact, the onset of disturbance in long distance telephonic communication—which is caused by sudden variations in earth currents, may be taken as precursor of auroral display. Störmer's auroral stations in Norway are regularly kept informed by the *Telegraph Department* of sudden commencement of disturbance in line communication so that the stations may be ready for making auroral observations [32].

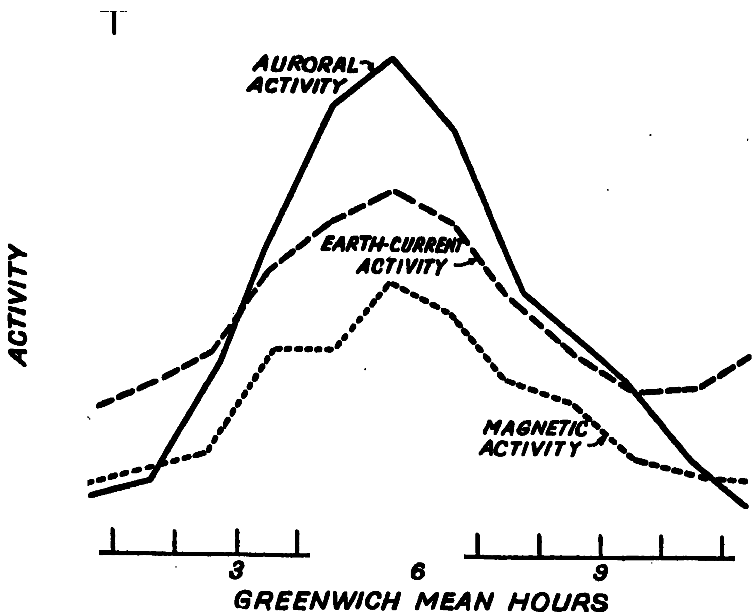


FIG. 12. Illustrating the similarity between the magnetic and the earth current activity on the one hand and the auroral activity on the other. (After Currie and Edwards.)

An interesting correlation has been established between the occurrence of the Types A and B of red auroras [see Sub-sec. (*i*)] and the spottedness of the sun. Type A occurs frequently in the years of maximum and Type B in the years of minimum solar activity.

(*l*) Intensity and colour of auroras

The intensity of auroras is generally estimated visually and expressed in terms of an international coefficient. A faint aurora, about as bright as the milky way is designated as having intensity I; one having a brightness of the order of a thin moonlit cirrus cloud is of intensity II; auroras as bright as moonlit cumulus cloud are of intensity III; and those emitting enough light to approximate moon light are of intensity IV. Thus, the illumination due to an arc of intensity III corresponds to that of a few microcandles per square centimetre and is about a hundred times that of the night air-glow which is about 10^{-8} candle per square centimetre [33] [Chapter X, Sec. 1].

A better and objective method of determining the intensity is to measure the total luminosity by means of a photoelectric cell as discussed in Sec. 2(*f*). It has been found by this method that the ratio of the luminosity between the brightest and the weakest auroras is as much as 1000 : 1 [34]. One great advantage in the use of the photoelectric cell is that measurements may be made even in foggy weather. (For a description of the method of using photoelectric cell the reader is referred to Chapter IV, Sec. 3*c*). Auroras with intensities I, II and III do not appear coloured, because their intensity is below the threshold of colour perception. For these auroras, the eye is affected only by the radiation λ 5577 which coincides with the maximum sensitiveness of the retina. Auroras with intensity IV appear coloured. They may be green due to λ5577 and violet due mainly to the second positive bands of N_2.

Auroras with red colour are specially interesting. The red colour may be due either to the forbidden radiation λ6300 of atomic oxygen, or to the first positive bands of N_2. The two types of red auroras may be distinguished by their characteristic colour distribution and are designated as Types A and B respectively.

Type A comprises the rays and streamers and the red colour extends throughout their whole length. Type B comprises the arcs and draperies, the red colour being confined only near the lowest edges. It has been found that red auroras of Type B may reach the unusually low height of 65–70 km.

The sunlit auroras, to which reference will be made presently, are mostly greyish-violet. On rare occasions they may also be blue. Their colour is thus distinctly different from the colour of ordinary auroras which, as already mentioned, is generally yellowish green and on some occasions deep red. The greyish violet colour of sunlit aurora is due to the red radiation of atomic oxygen mixed in large proportion with the green radiation of the same gas. The blue colour is mainly due to the enhancement of the negative bands of nitrogen.

(*j*) Sunlit auroras

The sunlit auroras, so called from the fact that they appear at great heights in the sunlit portion of the atmosphere, have been studied extensively in recent years. In Fig. 13 the group of auroral rays in the upper part depicts the mean positions of the rays in sunlit region [32]. The horizontal line divides the region of darkness below from that illuminated above. The small circles indicate the positions of the points measured in the auroral rays. Two such circles are plotted for each and the straight line joining them represents the auroral ray.

An interesting fact to be noticed in the figure is that some of the rays lie partly in the shadow and partly in the illuminated region. The part of the ray coinciding with the line separating the region of darkness from that of sunlight is invisible [35].

Compared with the rays of ordinary auroras those of sunlit ones attain enormous heights and have been found to extend up to 1100 km. The lowest points of these rays are also at much higher altitudes than the lowest limits (viz., about 80 km.) of the ordinary auroras lying wholly in the

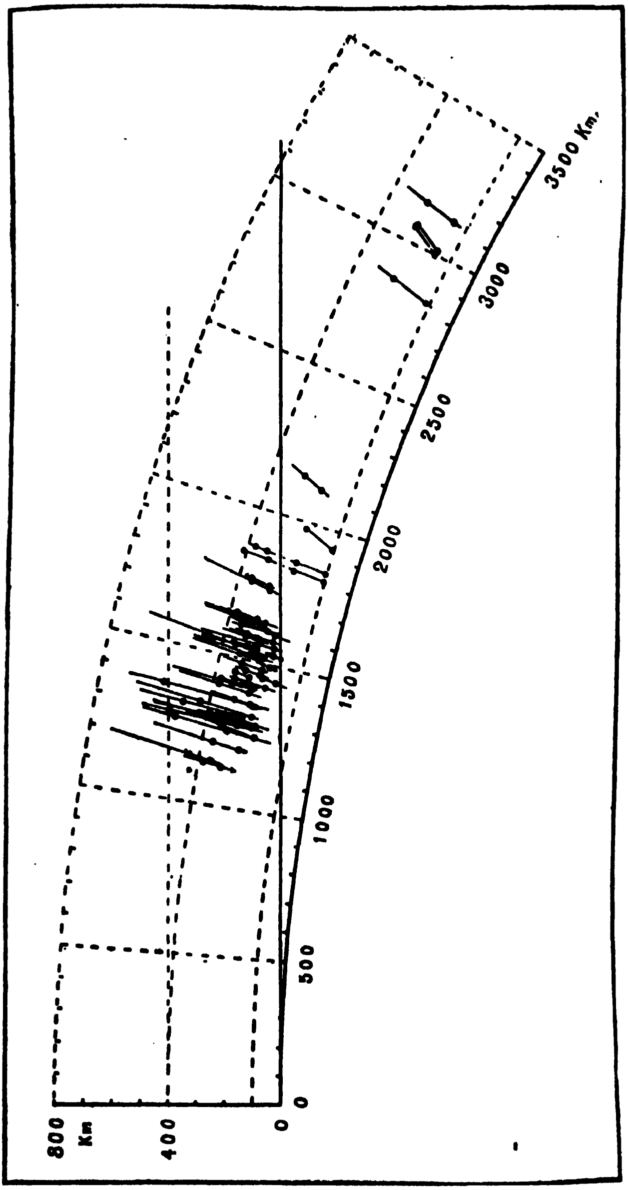


FIG. 13. Position of sunlit auroras compared with that of the shadow of the earth. The tangent to the earth's surface divides the region of darkness below from that illuminated above. The small circles indicate the positions of the measured points in the auroral rays. It may be noticed that some of the rays lie partly in the illuminated and partly in the shadow region. (Observed at Norway on March 15, 1929, by Störmer.)

shadow of the earth. Further, it will be noticed in Fig. 13 that the bottom edges of the rays lying wholly in sunshine have a tendency to follow the shadow-line of the earth.

3. AURORAL SPECTRUM

(a) The spectrum

We owe our knowledge of the auroral spectrum—the wavelengths of the lines and bands, their identification and intensity distribution—mostly to the systematic and long continued work of the Norwegian group of investigators—Störmer, Vegard, Harang and their collaborators. A valuable summary of the work up to 1945 is to be found in a communication by Vegard and Kvifte [36]. Fig. 14 shows typical auroral spectrographs and

FIG. 14. Spectrographs used for photographing auroral spectrum. (A) With adjustable glass optical systems. (B) With grating of high light-power for use in infrared region. The spectrographs were constructed at the new Auroral Observatory at Tromsö by L. Harang and E. Tönsberg.

Fig. 15, Plate II (facing p. 409) a typical auroral spectrum taken at the auroral observatory at Oslo [36a].

The lines and bands as are reported to have been observed in the auroral spectrum are those due to, (1) atomic oxygen, (2) atomic nitrogen, (3) neutral molecular nitrogen, (4) first negative bands of O_2^+ [36b], (5) first negative

bands and a new band system of N_2^+ discovered by Meinel [36c] and (6) occasionally lines due to sodium and hydrogen and probably also normal and ionized helium [37]. Presence of several emission lines from O^+ and N^+ have also been announced. These latter identifications are, however, still considered as doubtful. According to Nicolet most of these are to be ascribed to the one or the other N_2 band systems [38]. The presence of hydrogen line is very significant. When observed in the direction of the lines of magnetic force (terrestrial) the line ($H\alpha$) is found to be asymmetrically broadened towards the violet. This indicates that during auroral displays, hydrogen atoms (or, protons) are entering into the terrestrial atmosphere with high speed (see Chapter IX, Sec. 2*d*).

In Table III, the lines and bands which are more or less always present in auroral displays and whose identifications are considered as satisfactory are listed. It will be seen that for atomic oxygen there are strong permitted lines in the infra-red, besides the familiar forbidden red and green lines which are also observed in the night air-glow. For atomic nitrogen also, forbidden lines, involving the lowest metastable levels as in the case of oxygen atom, have been observed [38*a*]. Further, there are also lines in the infra-red due to allowed transitions. These allowed transitions for O and N atoms were first identified by Meinel [38*b*]. These are the lowest transitions which may be observed above the forbidden transitions noted above. Strong emission of the N_2 first positive bands (2, 0) in the infra-red ($\lambda\lambda$ 7688, 7717, 7746) has also been identified by Meinel [38c].

It is interesting to note that emission in the range of micro-waves have been reported by some workers [38*d*].

(*b*) Spectral characteristics of auroras: The type, height and latitude effects

The spectral composition and the intensity distribution vary considerably from one aurora to another depending on the type, the altitude and the latitude of the place of observation. In Fig. 16 the average distribution of intensity in the lines and bands are shown after Vegard [39]. In what follows we shall give short accounts of these variations characterizing them as the *type* effect, the *altitude* effect and the *latitude* effect.

Type effect.—The auroras, according to their colours, may be classified, independently of their shapes, into three types: (i) the ordinary type having a yellow-green colour, (ii) the *A*-type in which the red colour extends throughout the whole length (rays and streamers), and (iii) the *B*-type in which the red colour is confined to the bottom (arcs and draperies).

As already mentioned the yellow-green colour is due to the auroral green line of atomic oxygen. The red colour of the *A*-type is due to the red oxygen lines (λ 6300) and that of the *B*-type to the first positive bands of nitrogen. The following further remarks may be made about the spectral composition depending on the type of auroras—diffuse forms and those having distinct forms [40]:

TABLE III

Spectral Composition of auroras

Origin	Transition	Remarks
O atom ..	$^1S \longrightarrow {}^1D$ (λ 5577)	Always present in auroral spectrum—commonly called 'the auroral green line'—generally of the maximum intensity. Transition forbidden.
	$^1D \longrightarrow {}^3P$ (λ 6300, 6363, 6392)	Considerably intense in high and/or sunlit rays. Transition forbidden.
	$3^5P \longrightarrow 3^5S$ (λ 7774) $3^3P \longrightarrow 3^3S$ (λ 8446)	Allowed transitions. Strong emission.
N atom ..	$^2P^0 \longrightarrow {}^4S^0$ (λ 3466)	Ordinary aurora. Transition forbidden.
	$^2D^0 \longrightarrow {}^4S^0$ (λ 5198·5, 5200·7)	Low latitude, high altitude auroras. Transition highly forbidden. The radiation may be observed hours after the auroral activity has ceased.
	$3p^4D \longrightarrow 3s^4P$ (λ 8684, 8721)	Allowed transitions. Strong emission.
O_2^+ ion ..	$^4\Sigma \longrightarrow {}^4\Pi$ (Visible region.)	First negative bands. All bands of this system are rather weak.
N_2^+ ion ..	$A'^2\Sigma \longrightarrow X'^2\Sigma$ (Blue-violet and ultra-violet region.)	First negative bands; strongest of all nitrogen radiations. The intensity distribution is similar to that observed in laboratory.
	$A'^2\Pi \longrightarrow X'^2\Sigma$ (Infra-red.)	New band system discovered by Meinel; several of the bands are as intense as the first positive system of N_2.
N_2 molecule ..	$B\,^3\Pi \longrightarrow A\,^3\Sigma$ (Red and infra-red regions.)	First positive bands of N_2; most numerous in the spectrum.
	$C\,^3\Pi \longrightarrow B\,^3\Pi$ (Ultraviolet region.)	Second positive bands of N_2; the intensity is low.
	$A\,^3\Sigma \longrightarrow X\,^1\Sigma$ (Blue-violet and ultra-violet region.)	Vegard-Kaplan bands. Very weak in the auroral spectrum. Transition forbidden.
H-atom ..	H_α line (red) (λ 6563)	Broadened towards ultra-violet, indicating that fast H-atoms (or protons) are entering the terrestrial atmosphere.

(i) The intensity of the green line relative to that of the negative nitrogen bands is smaller for diffuse forms like grey and pulsating surfaces than for the sharper forms like bands and draperies.

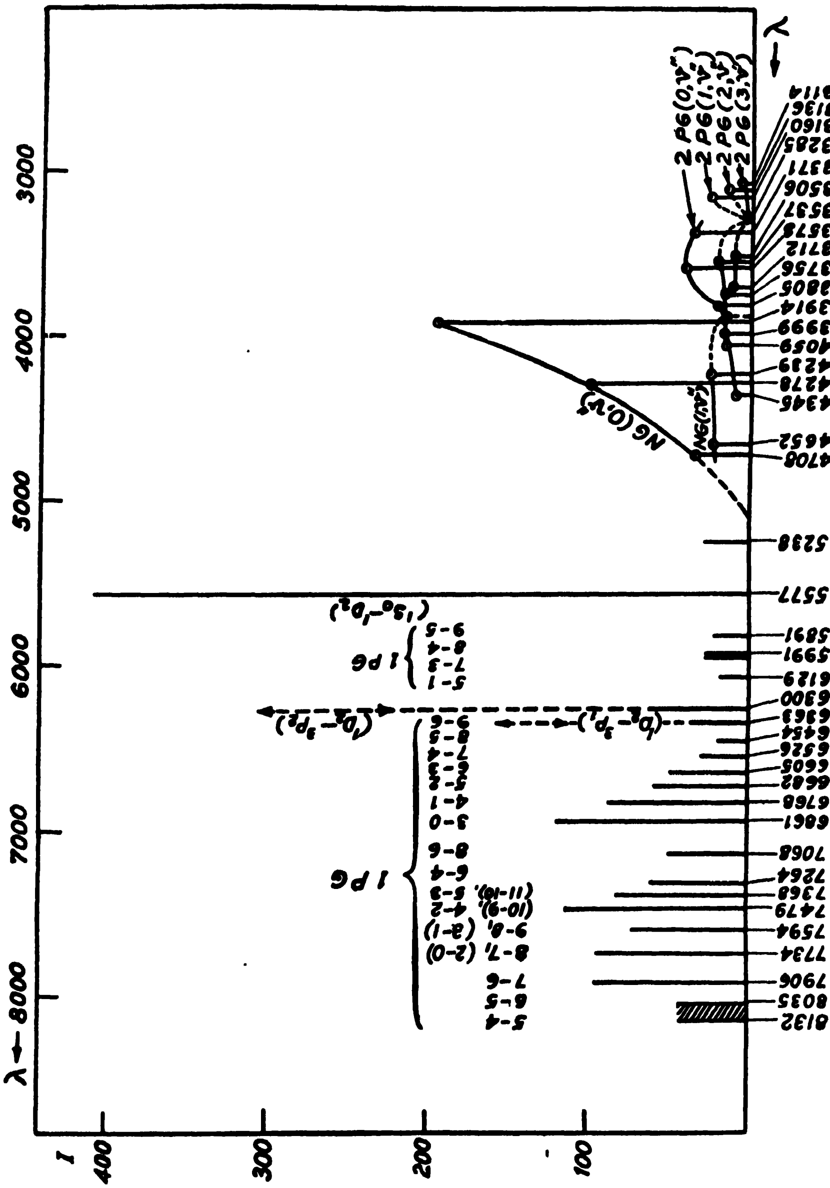


FIG. 16. Typical energy distribution in auroral spectra. (After Vegard.)

(ii) The intensities of oxygen red lines and of the first positive group of nitrogen are greater for the former than for the latter.

(iii) Some lines and bands in the blue region of the spectrum are more prominent in some of the diffuse auroras.

The change of spectral composition from the distinct to the diffuse forms is of the same nature as the change observed with increase of altitude. It may thus be thought that the difference between the two groups is merely an altitude effect, i.e., the average height of the diffuse form is greater than that of the distinct form. Actual measurements of height show, however, that there is no marked difference between the average heights of the two groups. It is, therefore, concluded that the two groups of auroras belong to two distinct spectral types.

For the sunlit auroras two types of intensity distribution may be distinguished. One in which the red oxygen lines are greatly enhanced [41] and the colour of the aurora is greyish violet and the other in which the negative bands of N_2^+ appear with great intensity and is of a bluish colour [42].

Height effect.—It has bee n found that in the same aurora, e.g. a streamer the ratio of the intensities of two spectral components changes with height. The main results observed are as follows [43]:

The intensity (I) of the red oxygen radiation increases relatively to that of the green radiation with height. The ratio $(I_{6300}/I_{5377})_{\text{upper}}:(I_{6300}/I_{5377})_{\text{lower}}$ varies between 1·4 and 4·25, the average value being 2·6. The height difference between 'lower' and 'upper' vary between 60–100 km.

The intensity of the N_2^+ negative bands increases relatively to that of the green oxygen line with height. The ratio $(I_{4278}/I_{5377})_{\text{upper}}:(I_{4278}/I_{5377})_{\text{lower}}$ varies between 1·4 and 1·9. The altitude variation is the same as before.

The intensity of the Vegard-Kaplan bands of N_2 increases relatively to that of the N_2^+ negative bands with height.

The intensity of atomic nitrogen lines increases relatively to that of the N_2^+ negative bands with height.

The study of these height effects is of great importance in determining the variations of composition in and the physical state of the upper atmospheric regions. Further observations are, however, necessary before these results can be properly interpreted and fully utilized.

Latitude effect.—The auroral characteristics change markedly with latitude.

It has been observed that the intensity of the red oxygen lines and the frequency of the red auroras of type A increase as one passes from the auroral zone towards lower latitude [44]. (The comparison was made during the same period of observations at Tromsö and at Oslo.)

It appears that the mean altitude of the auroras increases as one moves from the auroral zone towards lower latitudes [45]. Such high altitude auroras are characterized by the λ 5200 line of atomic nitrogen and a high intensity of the red oxygen lines (λ 6300) relative to that of the green

line (λ 5577). This is illustrated in Fig. 17, Plate II (facing p. 409) after Dufay and Tcheng Mao-Lin. The picture is an enlargement of a spectrogram obtained on September 18, 1941 at the St. Michel Observatory, Haute Provence, France [46]. The high intensity of the red oxygen lines and the great prominence of the atomic nitrogen line are clearly seen. The intensity effect may, however, be *an altitude effect* (*vide supra*) rather than a latitude effect.

4. PASSAGE OF SWIFT PARTICLES THROUGH THE ATMOSPHERE

We now proceed to discuss the possible variations of intensity of auroras with height on the assumption that they are caused by bombardment of the upper atmospheric gases by charged particles—ions or electrons. But before doing so it would be useful to recall the data, relevant to this theory, which have been obtained in laboratory experiments on bombardment of gases by charged particles. Such experiments are on the determination of the efficiencies of ionization and excitation by the passage of swift particles, on the cross-sections of the atoms and molecules for these various processes when exposed to a beam of charged particles and the Wilson chamber experiments for determining the ranges of the charged particles. These laboratory experiments differ from the actual bombardment process in the upper atmosphere in two important respects. In the laboratory experiment the gas bombarded, whether it is in a Wilson chamber or in a discharge tube, has the same density throughout. In the upper atmosphere the density is variable, decreasing from bottom upwards. Secondly, the upper atmospheric phenomena are on a scale of entirely different magnitude from that of the laboratory experiments. For instance, in the case of laboratory experiments the lengths of the ionization track in the gas might be a few centimetres, whereas those in the upper atmosphere for auroral rays might be hundreds of kilometres.

(*a*) Probabilities of ionization and excitation by fast particles

Experiments on ionization by bombardment of charged particles show that for fast electrons the probability of ionization is small. The probability increases as the electrons slow down, and attains a maximum at about 100 electron-volt. It then decreases again. The total number of ions produced by a fast electron as it runs itself dead (in air) is about 30 times the number of kilovolts expressing the initial kinetic energy [47].

For heavier positive particles, Wilson chamber experiments show that high speed particles produce copious ionization very much more intense than electrons. In fact, while the tracks of positive particles in Wilson chamber appear as thick continuous lines, being densely packed by droplets due to copious ion formation, those for electrons appear broken, the individual droplets being sensibly apart and recognizable as such. As in the case of electrons, the probability of ionization by fast positive particles increases with decrease of speed; it attains a maximum at about the same velocity as that

of the electrons, namely electron-velocity corresponding to 100 electron volts [48]. This, of course, refers to α-particles in air, but the value probably does not differ very much for other positive particles. The total number of ions produced by the average α-particle of range 6·97 cm. is $2{\cdot}20 \times 10^5$ [49]. The number of ions produced per mm. of path increases as the range (or velocity) of the particle diminishes and attains a maximum for particles with about 3 mm. range when it is 6000 per mm. Afterwards it diminishes. For 2·1 cm. range it is 4500 per mm. It may be noted that the energy spent for producing an ion-pair is 32·3 volts for oxygen molecule and 35 volts for nitrogen molecule.

Experiments on the likelihood of excitation of atoms and molecules by direct bombardment of fast electrons show that the probability of such excitation (i.e. excitation without ionization) is extremely small. The likelihood is maximum when the electron energy is a little above that corresponding to the excitation potential and even then the effective cross-section of the particle is only a few per cent or a few pro-thousands of the gas kinetic cross-section.

For positive particles, observations show that the particles leave behind them a trail of excited atoms. Such excited atoms may be due, either to excitation by direct impact or, to the positive ions formed dropping into an excited state, instead of to the normal state, when neutralized by recombination with an electron. But in any case the observed luminosity is extremely feeble compared to that produced in discharge tubes.

(*b*) Range of fast particles

A question of considerable importance in the theory of aurora is the depth to which swift particles (charged or uncharged) coming from outside will penetrate into the atmosphere.

From Table I it is seen that the average lower limit of aurora is round 100 km. Auroras are also known to occur occasionally at heights as low as 65 km. The amounts of air above these heights, reduced to S.T.P., are about 1 cm. and 100 cm. respectively. The speed and energy necessary for such ranges are very high indeed. Unfortunately, we have no well established theoretical formula to estimate these quantities. One has to rely more or less on experimental data and on empirical relations.

For α-particles of radioactive origin accurate laboratory data are available. Thus, the range of α-particles from Thorium C, velocity $2{\cdot}063 \times 10^9$ cm./s. is 8·711 cm. of air at 15°C. and standard pressure; range of the same particles from radium, velocity $1{\cdot}511 \times 10^9$ cm./s. is 3·389 cm. [52]. There is an empirical formula for the range given by $R\rho_0 = K_v{}^{3/2}$ where v is the speed and K a constant characteristic of the particle concerned. According to this formula $K = 4{\cdot}2 \times 10^{-17}$ for α-particles and $2{\cdot}36 \times 10^{-17}$ for protons [50].

Data for Ca^+ ions and protons are of special interest as evidences of the presence of these ions between the earth and the sun have been obtained during magnetic storms [51, 51*a*] (see Chapter IX, Sec. 2*d*). The same order

of speed is deduced for both from the fact that auroras and, simultaneously, magnetic disturbances occur about a day (e.g. three-quarters to two or three days) after the appearance of bright spots on the sun, if it is supposed that the solar corpuscles producing these phenomena originate from such spots. No laboratory experiments have been made to determine the range of Ca^+ ions of such speeds. But, from comparison with the range of recoil atoms in argon it is concluded that such particles will have a range of only 0·15 cm. in air at standard pressure (instead of 1 cm. as required).

From what has been said above it is clear that particles of the order of 1000 km./s. speed (which is the speed obtained from spectroscopic and also from the 'delay' evidence as mentioned above) cannot penetrate to the observed auroral heights. It thus appears that streams of solar corpuscles travelling with such speed are accelerated, in some yet unknown way, to the requisite penetrative speed as they near the earth. This is not possible for a single charged particle. It may, however, be so for particles of a neutral ionized beam (see Chap. IX, Sec. 4*a*).

For electrons the range can be measured from the tracks of the electrons shot off when a pencil of homogeneous X-ray is passed through a gas. The initial energy of the electron is calculated from the X-ray data. For an electron of energy 58,000 volts the observed range is found to be 5·7 cm. The ranges may also be calculated after a theory developed by Bohr, but they are found to be only 60 to 65 per cent of the observed ranges [52*a*]. Lenard had attempted to measure the range of an electron-beam on the assumption of mass absorption by the gas traversed. Thus

$$I = I_0 \exp(-\mu m)$$

where I_0 and I are the original and the final intensity of the beam, μ the mass absorption coefficient and m the mass traversed per unit area across. We might define the range of the beam as the length of the path in traversing which there will be, say, 95 per cent absorption. The mass absorption coefficient varies with the energy of the electron, being $1{\cdot}8 \times 10^7$ for 25-volt electron and 13 for 339000-volt electron [53].

5. THEORETICAL CALCULATIONS OF INTENSITY VARIATION WITH HEIGHT

We are now in a position to discuss—if not to explain—some of the characteristic variations of auroral light intensity as depicted in Fig. 7. Curves I, II and III in the figure are what one would expect if light emission were associated with ionization and the high atmosphere were isothermal with the density decreasing exponentially with height. Curve IV depicting the intensity distribution in streamers, in which the intensity remains more or less constant with height is, however, unexpected.

(*a*) Positive particle bombardment

If the bombarding particles are positive ones the light emission may be assumed to be proportional to the ionization produced. The density

of ionization produced will, at any point, depend upon two factors—the density of the air at the point and the speed of the particle. Since the density of the air increases as the particle comes down the ionization will also increase. Further, as the speed of the particle diminishes the production of ions may increase or decrease according as the velocity of the particle on entry into the atmosphere had been greater or less than the critical velocity for maximum ionization ($v = 5{\cdot}94 \times 10^8$ cm./sec.). If the atmosphere be isothermal, the rate of increase due to the exponential increase of air density is much greater than the possible decrease effect mentioned above, and the ionization therefore increases along the path of the trajectory till the particle is near the end of its range. Curves in Fig. 18 depict the intensity distributions calculated on the above basis, for α-particles

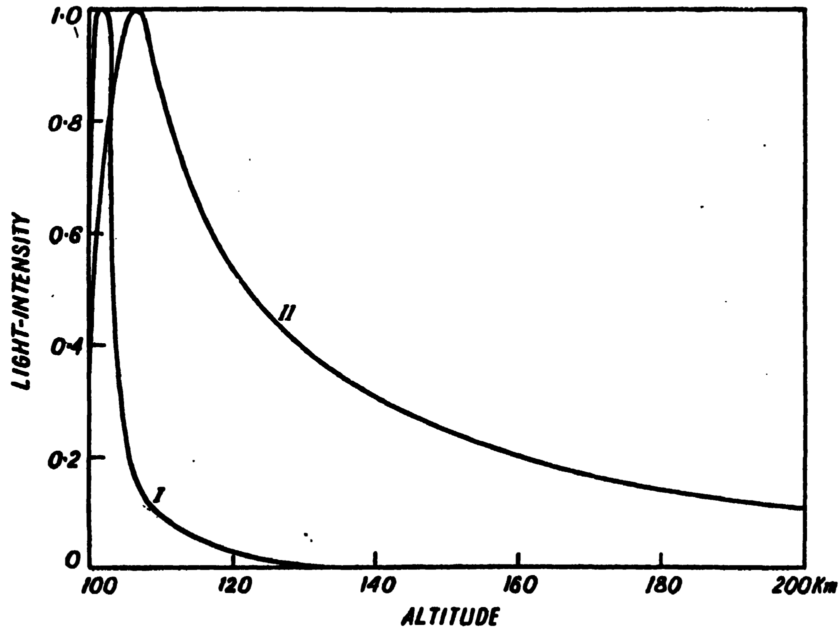


FIG. 18. Values of ionization (calculated) produced at different points of the trajectory of fast positive particles penetrating into the atmosphere. The ordinate scale is proportional to the intensity of emission which is taken to be proportional to the intensity of ionization. Curve I is for α-rays; Curve II for a heterogeneous bundle of canal rays. It may be noticed that the distribution of intensity resembles the curves of Fig. 7. (After Vegard.)

and for homogeneous bundles of other kinds of positive particles, by making certain assumptions regarding the density distribution [54]. It will be seen that the intensity distributions of these curves resemble those in Fig. 7.

(*b*) Electron bombardment

If the bombarding particles be electrons the intensity distribution, instead of being assumed to be proportional to ionization, may be taken to be

proportional to the *rate of absorption* of the electron rays along their trajectories. For obtaining the illumination curve, an auxiliary curve is first drawn, after Lenard, with the help of the formula for mass absorption, depicting the variation of intensity of the electron ray as it penetrates into the atmosphere. Another curve giving the *rate of change* of the electron-ray intensity at various points along the path is then plotted. This latter curve may be supposed to give light intensity distribution with height when the upper atmosphere is traversed by electron rays (Fig. 19) [54]. These curves are similar to the Curves I, II and III in Fig. 7.

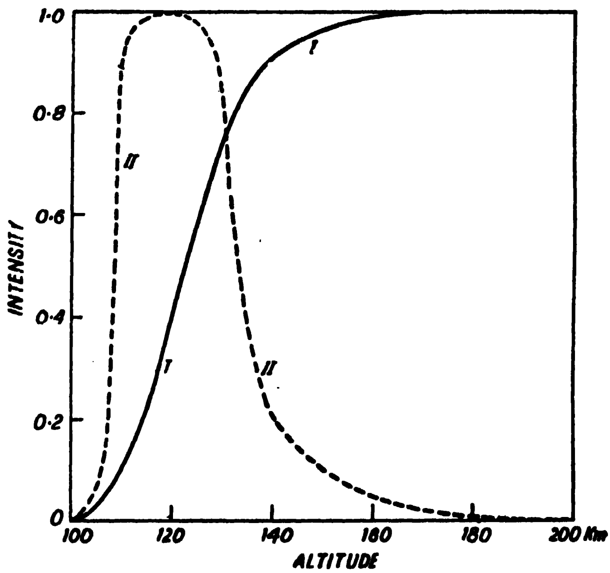


FIG. 19. Curve I—Variation of intensity of electron rays as they penetrate the atmosphere; the atmospheric density is assumed to vary exponentially with height. Curve II—Intensity of the luminescence produced by the absorption of electron rays. (After Vegard.)

The remarkable distribution of light intensity in auroral streamers (Curve IV, Fig. 7), namely a nearly uniform luminosity along the length of the streamer which may extend to some hundreds of kilometres, is difficult to explain. Several hypotheses have been put forward in this connection. It may, for instance, be supposed that the incident bundle of electric rays is composed of particles of different masses and velocities. Such particles, whether they follow straight paths along, or spiral paths round the lines of force [see Appendix, Sec. 7], will be absorbed at different heights. It may be imagined that the composition of the ray-bundle is such that the particles absorbed at different heights produce the same degree of ionization, that is, sensibly the same luminosity. Against this hypothesis it might be said that according to Störmer's calculations (see Chapter IX, Sec. 3b) the component particles of such a heterogeneous beam, coming from a great

distance, will enter the earth in widely different places and cannot follow the same line of force.

Another hypothesis based on the heating effect of the upper atmosphere by bombardment of charged particles, which does not seem to have been much discussed, may also be considered. Radio observations indicate that the atmosphere above 100 km. is far from isothermal. Indeed, available evidence shows that there is, on the average, a rising gradient of 4° K per km. above this level [see Chap. XI, Sec. 2 and Chap. XIII, Fig. 1]. This will cause a distribution of density with height falling much less rapidly than in the case of an isothermal atmosphere. Radio observations also show that, due to bombardment by fast charged particles there may be a further increase of temperature during magnetic storms. During auroral displays, which generally accompany magnetic disturbances, there may, therefore, be a large enhancement of temperature tending to make the distribution of density with height more uniform. The uniform illumination in auroral streamers may thus be a consequence of uniform density distribution due to heating caused by the impinging particles.

We may, in connection with the above hypotheses, discuss the possible causes of various auroral forms and their changes from one type to the other. Not infrequently one type of aurora gradually changes into another type—a homogeneous arc may gradually change over to a drapery-shaped arc and then transform into draperies and ray-bundles. The hypothesis that the different auroral forms are caused by bombarding electric rays gradually changing in composition—yet arriving at the same place of the earth is to be discarded for the reason already given. It is, however, possible that the variation from one type to another is due to a gradual variation of air density distribution within the auroral region. According to Vegard, however, such changes may explain the difference shown in different localities, and at different times in the same locality but cannot possibly account for the very large changes in intensity and type occurring in the same locality at about the same time. Vegard [55] is inclined to believe that the different types, as also their changes, are due to the ray bundles being composed of orbits, not parallel but meeting the lines of force at different angles. He has assumed various distributions-in-angle of the component trajectories, and has shown that the calculated intensity distribution given by these conforms to the observed distribution in the arc, the drapery and the streamers.

(c) Spiral trajectories and auroral streamers

Finally, we shall discuss the intensity distribution due to the so-called 'spiralling' trajectories. In our discussion so far we have tacitly assumed that the ionizing bundle of rays—whether of positive particles or of electrons—move approximately in straight paths following the lines of force. These trajectories, if they were able to proceed ideally, would reach the centre of the earth-magnet. The spiralling trajectories, however,

approach close to the earth along a spiral path and then recede after coming to a minimum distance.

The distance d of the turning point from the pole is, as shown in Appendix, Sec. 7, equal to $r \sin \psi$, where ψ is the angle which the trajectory makes with the lines of force on entry and r the distance of the point of entry from the pole. If the particle slows down due to energy spent in ionization then the number of turns would increase but the height of the turning point would remain unaltered. If ψ is large, approaching 90°, the number of turns will be large and the particle, as it will have to traverse a much greater length of air, might run itself dead before reaching the turning point. The ionization, as measured along the length of the lines of force, will obviously be greater particularly near the turning point of the trajectory.

Consider a trajectory which is completely absorbed *before* reaching the turning point. Let the angle which the trajectory makes with the lines of force (supposed to be nearly vertical) at the height h, where large absorption is taking place, be ψ_h. Then, if ψ_h be regarded as constant in the small interval of path in which the predominant part of the absorption is taking place, the air-equivalent r_h at this height for vertical incidence, will have to be multiplied by $1/\cos \psi$ to give the effective air-equivalent for the trajectory under consideration [55a]. If the lines of force round which the trajectory spirals be not vertical, but have a mean inclination I, then the air-equivalent has to be multiplied by $1/\cos \psi \cos I$. Thus, if the penetrating power of a ray be R_0 cm. of air, the depth to which it will penetrate when moving (near the point of predominant absorption) round lines of force of inclination I and making an angle ψ_h with the same, will be that at which the air-equivalent r_h (that is, the amount of air mass above the point) is such that $r_h/\cos \psi \cos I = R_0$. A particle of a specified range will therefore penetrate different heights of the atmosphere depending on the angle which the trajectories make with the lines of force. In general the bottom-edge of the luminous track will be higher for the spiralling trajectories than for those which follow a straight course in the direction of the lines. For the latter the height of the bottom edge is fixed by the penetrating power of the ray. For the former, however, the lower edge of the track need not be so fixed. A particle with a high penetrating power may reach the turning point without running itself dead and then recede. And this height is always greater than the height limited by penetrating power. According to Störmer, Vegard and others the auroral streamers are produced by spiralling trajectories. It is found that the streamers have, in general, a higher bottom edge than other types of auroras [see Table I, p. 412].

6. EXCITATION OF AURORAL SPECTRUM

(a) Introduction—Excitation of the first negative bands of N_2

The primary source of excitation of the auroral spectrum in the upper regions of the atmosphere is undoubtedly bombardment by fast charged corpuscles of solar origin. Simple considerations show, however, that this

bombardment cannot be the immediate cause of excitation of *all* the observed lines and bands.

TABLE IV

Comparison of the spectral characteristics of aurora and of night air-glow.

	Aurora	Night air-glow
Geographical distribution	Observed in polar regions.	Observed all over the world.
Height of occurrence ..	Occurs most frequently in the region 90–120 km.; sunlit auroras occur at much greater heights—350–550 km. and occasionally extend up to 1100 km.	Luminescent layers located between 100–400 km. (Regions *E* and *F* of the ionosphere.)
Primary source of energy	Charged particles from the sun.	Ultraviolet radiation from the sun. (And also bombardment by particles from interplanetary space ?)
Conditions under which observations are made	Observations are made when the source is in action. In the case of sunlit aurora the ultra-violet solar rays are also acting.	Observations generally made at night when the primary source is not acting. In the special case of early morning and evening observations the source may be acting on the high atmosphere.
Spectra:		
(1) First negative bands due to $N_2^+ (A' \rightarrow X')$	Strongest of all nitrogen radiations. Greatly enhanced in sunlit aurora.	Generally absent; but when the high atmosphere is illuminated by the rays of the rising or the setting sun, they are very strong.
(2) First positive bands of $N_2(B \rightarrow A)$	Strong in ordinary aurora.	Strong.
(3) Second positive bands of $N_2(C \rightarrow B)$	Comparatively faint.	Very faint.
(4) Vegard-Kaplan bands of N_2 $(A \rightarrow X)$	Very weak.	Very strong.
(5) Forbidden lines of O λ 5577 (green) $(^1S \rightarrow {}^1D)$ $\left.\begin{array}{l}\lambda\ 6300\\ \lambda\ 6363\end{array}\right\}$ red $(^1D \rightarrow {}^3P)$	Green radiation is very strong, usually much stronger than the red. But in going up an auroral streamer the intensity of the red lines increases. In sunlit aurora the red is greatly enhanced. The intensity of the red is about 4-5 times that of the green.	Both green and red are very strong and are of nearly equal intensity. In high sunlit atmosphere the intensity of red lines is greatly increased while that of green is hardly affected.
(6) Forbidden lines of N	Observed in low latitude high altitude.	Not definitely identified.
(7) Allowed lines of O and N	Certain allowed lines are strongly emitted in Aurora. See Table III.	Not observed.
(8) H_α line	Occasionally present. See Table III.	Not observed.

Firstly, if the excitation were due to direct bombardment, the different spectral lines and band systems would develop at different heights. Experimental results show that for simple excitation (without ionization) the energy of the bombarding electrons must lie close to the exciting voltage. The bombarding particles as they penetrate into the atmosphere attain the required velocities at different heights within narrowly defined limits. The corresponding spectra should therefore be emitted only from these strata unless of course one assumes that the bombarding particles possess a wide range of velocities.

Secondly, the auroral spectrum has certain similarities with the night air-glow spectrum (see Table IV) and this latter is excited even when there is no bombardment by solar corpuscles.

It is therefore convenient to classify the auroral spectrum into two categories according to the mode of excitation: (i) spectra excited as a result of direct bombardment, and (ii) spectra excited as a result of reactions amongst the neutral and charged particles, e.g. electrons, positive ions produced by the bombardment.

Regarding the first, it is now generally agreed that the only part of the spectrum excited by direct bombardment is the first negative bands of nitrogen molecules thus,

$$N_2+e \rightarrow N_2^+(A')+e+e.$$

[*Note*:—The bombarding particle has been shown as an electron; but it may also be a positive ion (e.g. Ca^+ and/or H^+ ion) or a neutral particle, e.g. H atom.]

It will be noticed that as a result of the impact the N_2 molecule is ionized and at the same time, raised to the A' state from which the negative bands originate (see Fig. 17, Appendix, Sec. 2). The hypothesis that the negative bands are excited by direct bombardment, receives strong support from the fact that these bands are easily excited in discharge tubes (where the bombarding process is operative) and also that the intensity distribution is similar to that in the aurora.

Difficulty, however, arises in the interpretation of the other spectral components, e.g. the forbidden red and green lines of oxygen, the first and second negative bands of N_2 and also, how these particular lines and bands are excited while others are absent or extremely weak.

Many theories have been proposed to explain these characteristics of auroral spectrum. But none of them can be said to be entirely satisfactory. In what follows we shall give brief accounts of some of the important theories and discussions.

(*b*) Forbidden atomic oxygen lines and the N_2-bands

(i) *Vegard's theory.*—According to Vegard [55*b*] the N_2 band systems are all excited by direct electronic impact and that the main problem is to explain how only forbidden emissions and not the allowed ones of atomic oxygen (as then unknown) occur. For the explanation Vegard assumes the presence

of nitrogen molecules loaded with higher energy (E_a)—the so-called active nitrogen molecules. The assumption is not unreasonable since active nitrogen is known to be produced by strong discharge. The energy of the active N_2 molecule is assumed to be 9·55 eV. The following reaction with O_2 molecules is proposed,

$$E_a + O_2 \rightarrow O(^1S) + O(^3P) + \epsilon.$$

It is argued that since the surplus energy ϵ is only 0·25 eV; the reaction has a high probability. For the excitation of the red auroral lines the presence of ozone is further assumed. The probable reaction according to Vegard is as follows:

$$E_a + O_3 \rightarrow O(^1D) + \epsilon.$$

Vegard does not make any specific assumption regarding the nature of active nitrogen though, as mentioned above, he assumes its energy to be 9·55 eV.

The older views about active nitrogen are the following: (1) It is atomic nitrogen and that when the two atoms unite in a three body collision the energy of dissociation $D(N_2)$ is released (according Herzberg $D(N_2)$ is 7·36 eV; according to Gaydon it is 9·76 eV.). (2) It is molecular nitrogen in the metastable A-state from which the Vegard-Kaplan bands originate. According to more recent view of Mitra active nitrogen is N_2^+ (X') ions. (For an account of the theories and properties of active nitrogen, see Ref. [56].) The energy available in this case, in the neutralization of N_2^+ ion in the presence of a third body is 15·58 eV. Bates, Massey and Pearse [57] have examined the probabilities of emission with Vegard's reaction in which either the N atom or the N_2^+ ion is the active agent. In both the cases the emissions are found to be too weak. According to these authors the older theory in which $N_2(A)$ molecules is regarded as the active substance has more likelihood of being the energetic molecule of Vegard. The energy loading in this case is 6·15 eV (see Fig. 17, Chapter X). However, on various grounds, the authors attach little significance to the rôle of such molecules in the excitation of the auroral spectrum.

(ii) *Mitra's hypothesis.*—According to this theory, the first negative bands are excited by direct impact of charged solar corpuscles as already explained at the beginning of this section.

For the excitation of the first positive bands and the forbidden lines of atomic oxygen Mitra favours the same process as advanced by him earlier to explain the simultaneous emissions of these bands and lines in the night sky [58]. Mitra was led to this hypothesis by the existence of certain similarities between the auroral spectrum on the one hand and the night air-glow spectrum on the other, particularly when the high atmosphere is observed during twilight (see Table IV). The process is fully discussed in Chapter X, Sec. 2(a) and is only briefly noticed here.

Since oxygen atoms have very high electron affinity, O^- ions are readily formed. These O^- ions react with N_2^+ ions as follows:

$$N_2^+(X') + O^- \rightarrow N_2(B) + O(^1S) \quad \ldots \quad \ldots \quad \ldots \quad (I)$$

The process is one of so-called electron transfer and may be expected to have a high probability when the resonance condition is satisfied [see Chapter VI, Sec. 12, Table V]. Since the ionization potential of N_2 is 15·58 eV and the electron affinity of atomic oxygen is 2·2 eV, the energy released on neutralization is 13·38 eV. This energy is taken up almost entirely by both N_2 and O, if, on neutralization, the former drops to the vibrational level $v' = 9$ of the B-state (9·1 eV) and the latter is excited to the ^1S state (4·2 eV). The N_2 (B) molecules thus produced come down to the A-state and emit the first positive bands and the O (^1S) atoms to the ^1D and then to the ^3P state and in the course emit the green and the red forbidden lines respectively [see Fig. 17, Chapter X]. If account is taken of the fact that some of the N_2 molecules will be in low vibrational states on account of the high temperature of the region (average temperature 1000°K corresponding to 0·2 eV) then to accommodate this extra energy the neutralized N_2 molecules will drop to the vibrational levels $v' = 10, 11$ or 12. The emissions corresponding to transitions from these levels are just those that are found enhanced in the first positive band system in the auroral and the night sky spectra. (There is some uncertainty about the value of the electron affinity of O. While Lozier [59] gives the value as 2·2 eV, recent experiments of Vier and Mayer [60] yield the value 3·0 eV. The former is, however, the generally accepted value and has only been taken into consideration here.)

Second positive and the Vegard-Kaplan bands.—Reaction (I), on account of the inadequacy of the energy, cannot raise the N_2 molecules to the C-state. As such, it cannot explain the emission of second positive bands of N_2. That the mode of emission of the second positive bands must be different is also obvious from the following consideration. The auroral spectrum shows that the N_2 molecules, after emission of the second positive bands, are left in the levels 1–4 of the B-state. But the most intense first positive bands originate in levels 11 to 8 of the same state.

The faint emission of second positive bands can be explained if it is assumed that some of the N_2^+ ions combine directly with electrons by the radiative process producing neutral molecules in the C-state. We thus have

$$N_2^+(X')+e \rightarrow N_2(C)+h\nu. \quad \ldots \quad \ldots \quad \ldots \quad \text{(II)}$$

The cross-section of N_2^+ for such reaction is 10^{-19} cm.^{2} [61] and is small compared to the cross-section for reaction (I) which is 10^{-14} cm.^{2} [62]. The presence of comparatively faint second positive bands both in the auroral and in the night sky light is thus explained.

As shown in Table IV, there is a marked difference in the intensity of the Vegard-Kaplan bands in the aurora and in the night air-glow. Since the N_2 molecules after emission of the first positive bands are left in the A-state, it might seem strange why the V-K bands which are due to transitions of the N_2 molecules from this state to the X (ground) state, would be weak in the former but strong in the latter. The explanation is, however, simple. The A-state of N_2 is known to be a highly metastable

state, i.e. it has a long life. Now the metastable molecule may come down to a lower state by either radiation of energy or by a radiationless process by giving up its energy to another particle with which it collides. In order that the process may be a radiative one it is necessary that the average interval of collision must be greater than $T\epsilon$ where T is the average lifetime of the metastable state and ϵ is the efficiency of collision. ($1/\epsilon$ is the average number of collisions necessary to bring the metastable particle down by a radiationless process.) Now, the auroral spectrum is emitted from heights which are low (90–120 km.) compared to those from which the O-lines and N_2-bands of night air-glow are emitted. The collisional frequency in this region being higher, the metastable $N_2(A)$ molecules are brought down to the normal state rather by collision with the other atmospheric particles than by radiation of the V-K bands. That the relative weakness of the V-K bands in the auroral light may be due to the higher collisional frequency at auroral levels has also been suggested by Chapman [63].

Criticisms of the theories.—Bates, Massey and Pearse [57] have closely examined the various theories put forward for the excitation of the auroral spectrum. They agree with Mitra that insofar as the first negative bands are concerned, these are excited as a result of direct bombardment by solar corpuscles. They, however, doubt the adequacy of the radiative recombination process for the second positive bands on the grounds that the recombination coefficient is too low for the purpose, and also that there is no reason why in the recombination the state N_2 $(C^3\Pi)$ should be preferred to the numerous other possibilities. In regard to the excitation by charge transfer process (I), they are of opinion that this need not be the only process of excitation. Instead, they consider that a number of mechanisms (including direct excitation and charge transfer) are involved. As there will be large increase in electron concentration (with electrons of energy up to 20 eV), it is quite likely that there will be excitations by electronic impacts also. They note in particular the possibility of the dissociative recombination process as follows:

$$O_2^+ + e \rightarrow O' + O''(\text{states } ^3P, {}^1D, {}^1S)$$

which they had suggested earlier to explain the forbidden atomic oxygen lines in the night sky. The relative importance of the different processes depend on factors such as electron energy, degree of ionization, the gas density and constitution, and, hence the wide spectral range as observed could be anticipated. However, as the present state of our knowledge of these various factors are still inadequate and as opposing influences are frequently involved it is not yet possible to explain the full complexity of the auroral spectrum.

(c) Forbidden atomic nitrogen lines

It has been proposed by Mitra [64] that dissociative recombination of N_2^+ ion, followed by production of N-atoms in the metastable ^2P and

^2D states can account for the forbidden atomic nitrogen lines in the aurora. Consider the reaction

$$N_2^+(X')+e \rightarrow N(^2P)+N(^2D)$$

assuming that the dissociation energy of N_2 is as given by Gaydon viz. 9·76 eV [65]. The energy supplied by the left-hand side is 15·58 eV (first ionization potential), the N_2^+ (X') ion being in the lowermost vibrational level. The energy demanded by the right-hand side is 15·69 eV being the total of those required for dissociating the neutral N_2 molecule (9·76 eV), for exciting one nitrogen atom to the ^2D state (3·56 eV) and for exciting the other atom to the ^2P state (2·37 eV). The demand is thus slightly in excess of supply. Now, the observed N_2^+ band systems show that the excited $N_2^+(A')$ ions, from which the bands originate, are left, after emission, not only in the lowermost vibrational level $v''=0$, but also the higher levels $v''=1$, 2, etc. of the $N_2^+(X')$ state. The dissociative recombination can thus take place with the $N_2^+(X')$ ions as are in the higher vibrational levels $v''=1$, 2 and are loaded with energy slightly in excess of that demanded by the reaction. [Alternatively, as has been pointed out by Bates (private communication), the kinetic energy of fast incident electrons may supply the slight deficiency in energy.] The $N_2^+(X')$ ions which are left in the lowest vibrational state $(v''=0)$ react with O^- to excite the lines and bands as indicated by Eq. (I) in sub-sec. (*b*) above. The slight defect in energy also explains the absence of atomic nitrogen lines in the night air-glow. Because, though electrons and $N_2^+(X')$ ions are both present in the upper atmospheric regions where the night air-glow originates, the ions, in the absence of any bombarding corpuscles, are mostly in the lowest vibrational state. It is to be noted that similar dissociative recombination for the case of O_2^+ ion is advocated by Bates and Massey to explain the emission of forbidden atomic oxygen lines.

If nitrogen atoms are already present in the upper atmospheric regions, then the emissions of the atomic N-lines may be assumed to be caused by direct electronic impacts. Processes have been suggested by which N-atoms may be produced as a result of extreme ultraviolet light absorption.

Thus, the vibrationally excited $N_2{}^+(X')$ ions as are necessary for the dissociative recombination process discussed above may be produced as a result of absorption of solar radiation $\lambda < 661$, by which the N_2-molecules are ionized and, at the same time, raised to A'-level (see Appendix, Sec. 2, Fig. 17). After the emission of the negative bands $(A' \rightarrow X')$, those of the $N_2{}^+(X')$ ions which are left in the higher vibrational levels undergo dissociative recombination and N-atoms are produced as explained above.

Again, according to Herzberg and Herzberg [65*a*] absorption in the wavelength range $\lambda\lambda$ 1150–1250 in the region of Lyman-Birge-Hopfield bands (see Appendix, Sec. 2, Fig. 17) may lead to predissociation—the predissociation limit being between the levels $v'=6$ and 7 of the $a\ ^1\Pi$ state. As the absorption is comparatively weak (the bands representing a forbidden transition) the radiation will penetrate deep and the formation of the

nitrogen atoms will go on over a considerable range of heights. This explains that notwithstanding the dissociation, N_2 molecular bands are observed from greater heights, e.g. sunlit aurora. [The discussion of the possibility of dissociation of N_2 molecules as a result of the extreme ultraviolet light absorption may be taken as a continuation of the same in Chapter V, Sec. 3.]

(*d*) Allowed atomic oxygen and nitrogen lines

No specific processes have been suggested for the excitations of the strong allowed lines of atomic oxygen and nitrogen as listed in Table III. It will be noticed that the energy of excitation is high, being 10·5 eV for the oxygen lines (see Fig. 16, Appendix, Sec. 2). One is tempted to suppose that the atoms are excited to these high energy states by the direct impact of the impinging fast corpuscles. But, as mentioned in the introduction, for such case the energy of the colliding electron should lie close to the exciting energy and the required speed of the bombarding corpuscle will be attained within narrow limits of heights. Further, as the intensities of these radiations are quite high, the hypothesis that the atoms are excited in the process of radiative recombination is not favoured; because, it is well known that the probability of radiative recombination is very small.

(*e*) The spectrum of sunlit auroras

The spectral characteristic of sunlit aurora consists in the enhancements of the red lines of the atomic oxygen (leaving the green line unaffected) and of the first negative bands due to N_2^+. These effects are strikingly analogous to those of the night sky spectrum during early morning or evening hours when the rays of the rising or the setting sun touch the high atmosphere (Fig. 20, Plate II, facing p. 409). These enhancements are simply explained as due to the direct action of the solar ultraviolet rays on the oxygen atoms and on the N_2 molecules.

It is to be noted that the spectra (*a*) and (*c*) were taken when the high atmosphere was illuminated by the solar rays and spectra (*b*) and (*d*) when the same was in the dark. For (*a*) the height of the sunlit auroras was 400–600 km. and the exposure from $2^h\ 35^m$ to $3^h\ 55^m$. For (*b*) the height was 90–120 km. and exposure from $21^h\ 45^m$ to $21^h\ 55^m$. The exposure for the twilight spectrum (*c*) was 15 min. only and for the night air-glow spectrum (*d*) 1 hr.

Taking the case of oxygen atoms we note that they are raised to ^1D state (from which the red lines are emitted) by absorption of $\lambda < 6300$ and to ^1S state (from which the green line is emitted) by absorption of $\lambda < 2972$. But the number of O atoms raised to the ^1D state is much larger than that raised to the ^1S state. This is due to two factors. Firstly, the absorption coefficient of O atom for excitation to the ^1D state is much higher than that for excitation to ^1S state, and secondly, the density of solar radiation in the

spectral region $\lambda < 2972$ corresponding to transition to the ^1S state is about 100 times smaller than that in the region $\lambda < 6300$ corresponding to transition to ^1D state. As a result of this, the red oxygen lines are emitted in much greater intensity than the green line. (These points are discussed in greater detail in Chapter X, Sec. 2*b*.)

Regarding the enhancement of the first negative bands of N_2^+ there are two possibilities.

(1) The N_2 molecules have strong absorption in the region $\lambda < 661$. Absorption of solar extreme ultraviolet radiation in this region ionizes and, at the same time, raises the ionized molecule to the A' state from which the negative bands originate (Fig. 17, Appendix, Sec. 2).

(2) N_2^+ ions in the ground (X') state are already present in the high sunlit regions and that by absorption of radiation in the region $\lambda < 4708$ the ions are raised to A' state from which the negative bands are emitted. Both these views are the same as those advanced to explain the 'twilight flash' in the upper atmosphere (see Chapter X, Sec. 7*b*).

(*f*) Height effects

The 'height effects' as described in Sec. 3(*b*) can generally be explained as due to de-activation by collision of the metastable atoms and molecules which contribute to the auroral spectrum, though, other causes may also be operative. We may, for example, examine the increase of intensity of red oxygen lines, relative to the green line, with height.

According to reaction (I) p. 438 the O-atoms are raised to ^1S state and then come down to the ground state by successive radiations of the green and the red lines $^1S \rightarrow {}^1D \rightarrow {}^3P$. Since for every emission of the green line there must be a corresponding emission of the red lines the intensities of the two should be of the same order. (The probability of emission by $^1S \rightarrow {}^3P$ jump is considered to be very small compared to $^1S \rightarrow {}^1D$ or $^1D \rightarrow {}^3P$.) The difficulty is removed if the great difference in the lives of the oxygen atoms in the metastable states ^1S and ^1D is recalled. The value of the former is 0·5 second while that of the latter is 100 seconds. At lower levels, on account of higher collisional frequency, a larger proportion of ^1D atoms comes down to the ground state by radiationless process. With increasing height the collision frequency decreases and an increasing number of atoms in the ^1D state has chance of emitting the red lines.

The intensity increase with height of the Vegard-Kaplan bands and of the atomic nitrogen lines (both of which originate from highly meta-stable states) can also be similarly explained as due to the decrease of collisional de-activation with height. The reader interested in quantitative estimation of the collision de-activation effect may consult Ref. [66]. (It should be noticed that the increase of the intensity ratio N_2^+ negative bands/oxygen green-line, with height cannot be simply explained as due to collisional de-activation effect.)

There has been some confusion regarding the height effect as discussed above with the sunlight effect discussed earlier, since the two effects are

similar, namely enhancements of the red oxygen lines and of the negative bands of N_2^+. While Störmer [67] believed that the characteristics of sunlit aurora as observed by him were true effect of sunlight, Vegard was of opinion that the effect was merely an extension of the height effect as observed by him [68]. There, however, need be no confusion; because, the portions of the aurora studied by Störmer were actually within the sunlit portion of the atmosphere and the enhancements are exactly similar to the effect of sunlight on high atmosphere as first observed by Slipher [69] and later confirmed by Elvey [70]. Further, the region observed is generally 300 to 600 km. and is much above the region in which Vegard observed his height effect.

CHAPTER IX

THEORIES OF MAGNETIC STORMS AND AURORAS

1. INTRODUCTION

It has been known for a long time that terrestrial magnetic disturbances and the auroral displays have certain essential features in common. We have already seen in Chapters VII and VIII that both occur with great intensity and frequency in zones round the magnetic axis poles of the earth; both are also found to be closely associated with solar activity. Their intensities and frequencies of occurrence wax and wane over the 11-year solar cycle. Any theory proposed in connection with the one ought, therefore, in its fundamentals, to hold in respect of the other.

The theory, for instance, should explain:

(a) The distribution of auroras in space, in particular the auroral belt and the height of the most frequent occurrence of auroras; it should also explain the various distinct forms which the auroral displays assume.

(b) The origin of magnetic disturbance, both of the world-wide type D_{st} and also of that related to the local time S_D. Of the latter, it should explain in particular the marked localization of the disturbance near the auroral belt. In other words, it should explain the origin of the various upper atmospheric current systems to which these disturbances are due. It should also explain such minor magnetic disturbances as magnetic pulsations.

(c) The periodicity of the auroral and magnetic phenomena as related to the sun, e.g., the diurnal and seasonal variations, the 27-day recurrence tendency and the variations following the 11-year solar cycle.

The fact that magnetic disturbances and auroral displays are concentrated round the axis poles has long led to the belief that both the phenomena are caused by high speed charged corpuscles guided towards these regions by the action of the terrestrial magnetic field.

These ideas received strong support from laboratory experiments with 'Terrella' as were first performed by Birkeland [1]. An iron sphere surrounded by a magnetizing coil was placed in the centre of a large vacuum-chamber. On exposing the sphere to cathode rays produced in the chamber it was found that with a certain degree of magnetization and certain velocity of the cathode rays, the latter would strike the sphere round the poles (Fig. 12).

Regarding the source of the charged corpuscles it was suggested by Goldstein, as early as 1881 [2] that the particles may be of solar origin.

Perhaps the simplest theory is to assume that the sun occasionally sends out such particles (of one sign only) in the form of a beam. The beam thus constitutes an electric current stream and, on approaching the earth, it is so deflected as to precipitate round the polar region of the earth. This theory has been developed in great detail by Birkeland [3, 4] and by Störmer [5, 6]. The theory has many attractive features but nevertheless it has to encounter a fundamental difficulty first pointed out by Schuster [7]. To produce the auroral (and magnetic storm) phenomena the number density of the particles in the arriving stream must be very great. But, if the stream contains charged particles of one sign only then the particles would be dispersed by mutual electrostatic repulsion long before the stream reaches the earth from the sun. To obviate this difficulty Lindemann [8] proposed that the stream as a whole may be electrically neutral, being composed of equal amounts of both positive and negative particles. This idea was taken up by Chapman and Ferraro [9, 10] and later by Alfvén [11]. The former authors explained by their theory only the initial and the main phases of magnetic storms. The theory has, however, been extended by Martyn [12] by making use of certain hydrodynamical analogies and it is shown that it can give very plausible explanations of almost all the features of the auroral and geomagnetic disturbance phenomena. It appears now that the Chapman-Ferraro-Martyn theory is the only surviving theory of auroras and magnetic storms.

In what follows we will first give a brief résumé of the various suggested processes by which charged particles may be emitted from the sun. Short accounts of the different theories of magnetic storms and auroras will then be given. These will include, besides the theories in which the charged particles are assumed to be of solar origin, also the 'ultraviolet light' theory in which the particles are assumed to be produced in the earth's upper atmosphere (by the action of ultraviolet radiation) and not in the sun.

2. EMISSION OF CHARGED PARTICLES FROM THE SUN

(a) Milne process

Of the various possible sources of emission on the sun those that have been first considered are, naturally enough, the solar prominences. As is well known these are of common occurrence and appear on the solar disc as bright or dark filaments when examined with the spectro-heliograph [see Appendix, Sec. 8]. They are, however, displayed in their full grandeur when they occur on the limb of the sun and then provide opportunity for studying the details of the motion of the matter in them. Measurements show, however, that the velocities with which matter is shot upwards in the solar prominences are small compared to those required for corpuscles causing auroras or magnetic storms. The mean velocity is 153 km./sec. with occasional maxima of 400 km./sec. The prominences, as such, cannot, therefore, provide particles necessary for explaining the auroras or the

magnetic storms. Nevertheless, a study of the process by which matter in the prominences is thrown to great heights (10,000 km. or more above the sun's disc) *has* suggested a mode by which high speed ions may be ejected from the sun. This we now proceed to discuss.

Calculations show that matter projected upward from the sun would soon fall back to the surface unless supported by some force acting upwards. In fact, prominences are often observed to shoot upwards in jerks in defiance of the law of gravity as if they were receiving from time to time impulses from below [13]. According to Milne [14] these impulses are caused by 'radiation pressure' the origin of which is, however, quite different from that of the familiar radiation pressure exerted by light falling on reflecting or absorbing bodies of sizes large compared with the wavelength. This pressure is essentially dependent on selective absorption by the atom placed in a unidirectional radiation field. Consider an atom in the chromosphere exposed to radiation from the photosphere below. Assume for simplicity that the atom can exist in two states, a normal one and an excited one caused by absorption of radiation to which it is exposed. The wavelength of the radiation absorbed will be that of the corresponding Fraunhofer absorption line and, as such, its intensity will be much less (about one-tenth) than that of the continuous spectrum on either side, being only that due to the residual intensity of the line. The atom on absorption will go over to the excited state and at the same time receive an upwardly directed momentum, since, on absorbing a quantum of radiation an atom absorbs also the momentum possessed by the quantum. As long as the atom remains excited it will not experience any further impulse because it does not absorb any further radiation. After some time (of the order of 10^{-9} sec.) it will radiate and return to its normal state and will again be in a position to absorb. It will then receive a fresh impulse upwards and the process is repeated. The atom, of course, experiences a recoil momentum when it emits the absorbed radiation, but the recoil being in a random direction, its average is zero. The integrated effect of the unidirectional impulses received during absorption is the same as that of radiation pressure and the absorbing atom is held in suspension when the downward acceleration due to gravity which is always acting, is balanced by the pressure. The atoms in the chromosphere thus remain suspended about a certain level above the surface of the sun. The equilibrium may be compared to that of small bits of paper fluttering and held in suspension in an upwardly directed stream of air sometimes falling slightly and sometimes ascending by the wind pressure.

Now consider an atom which, at the instant of absorption, was moving upward with a high velocity, i.e., a velocity large compared with the average velocity but small compared to the velocity to be presently discussed. There is always a certain percentage of such atoms if the velocity distribution in the chromospheric gas particles be of a Maxwellian type. Due to Doppler effect the wavelength of the radiation which such an outwardly moving atom will absorb selectively is less than what it would if it were at

rest. The intensity of the altered wavelength is relatively high because it will lie outside the central darkest part of Fraunhofer line in the violetward wing. The atom is thus subjected to much higher radiation pressure as it has to bear almost the full brunt of the radiation intensity in the continuous spectrum, and is accelerated outwards. As before, after a certain interval the atom radiates experiencing a recoil momentum at random —the average of which is zero—and is again in a position to absorb radiation of still shorter wavelength and receive a still higher impulse. The process goes on and the atom moves further and further away from the sun until the 'dilution' of radiation due to inverse square law reduces the acceleration to zero. The atom may thus get right outside the gravitational field of the sun with a large limiting velocity. It is obvious that the particles that may thus escape are just those that are held in suspension by the 'radiation pressure' (H, He, Ca^+, Sr^+ and Ti^+). Milne has studied in particular the case of the Ca^+ ions which are present abundantly in the high levels of the chromosphere and finds that the limiting velocity is of the order of 1600 km./sec.

It is to be noted that the particles projected may be neutral or ionized. In the former case the neutral particles composing the emitted beam may be ionized on their outward journey. The beam as a whole would still be neutral because the electrons and the positive ions would be moving with same speed in the beams. For the latter case, however, conditions will be different. Imagine that positive particles are being projected from a particular region of the photosphere. The outgoing particles will induce a negative charge in the region and cause electrons, which are abundantly present due to thermal ionization, to escape also. The electrons, however, will not have the same speed as the positive ions as they are not subject to the same intense radiation pressure. (The only 'pressure' which they experience is that due to Compton scattering, which is extremely small compared to that on atoms due to selective absorption.) According to Milne, therefore, the aggregate of charged particles, Ca^+ ions and electrons for instance, will not form a neutral cloud but will have a positively charged head and a negatively charged tail and the head will be moving faster than the tail. These considerations have important bearing on the neutral corpuscular beam theory of auroras and magnetic storms to be discussed presently [Sec. 4].

(*b*) Alfvén's process

It has been suggested by more than one author that high energy particles may be produced in sunspots, which are seats of intense magnetic field of strength several thousand gauss (see Appendix, Sec. 8). According to Alfvén, the rapid whirling motion in the sunspots (due to the magnetic field) must produce strong electric polarization. And, such polarizations can give rise to discharges which would accelerate charged particles up to energies of the order of magnitude 10^8 electron volts [15].

Alfvén further discusses how these charged particles would be expelled from the sun if (as originally supposed by Hale) there is, as in the case of the earth, a general magnetic field at the surface [16]. However, much of Alfvén's arguments loses its significance as later observations cast grave doubts on the existence of such a field [17]. Nevertheless, we shall briefly describe Alfvén's process (on the supposition that the sun has a general magnetic field) as it contains many interesting ideas.

Consider first the simple case when the charged particles (positives and negatives in equal number) are produced within a small region at the equator with velocities directed normally to the solar surface. The particles, on account of the assumed general magnetic field of the sun, will not be able to leave the surface of the sun (in the absence of the process described below) even if they have energies of the order 10^8 eV, but will describe orbits round the lines of force. If the magnetic field (north-south) were homogeneous, the orbits would be circles (of radius about 10^4 to 10^5 cm. for 10^8 eV particles). However, as the field is not homogeneous (decreasing outwards from the sun) the paths described are only nearly circular and the positives and the negatives will drift (with nearly circular motions) along lines perpendicular to the field and its gradient, that is, the positives westward and the negatives eastward. Due to this motion, which we may call the *inhomogeneity drift motion*, the region will be polarized, with positive space charge on the west and negative on the east. There will thus be an electric field (F) west to east. The strength of the field will evidently be determined by the equilibrium between the rate of space charge production by the drifts on the one hand and the rate of neutralization of these charges on the other.

Note.—The drift in an inhomogeneous field is understood as follows. Consider the motion of the particles in the equatorial plane of the sun. Remembering that the magnetic lines of force run perpendicular to the plane and that the force decreases outwards, the part of the (nearly) circular path close to the surface will be in a stronger field than the part remote from the same. As such, the former part of the path will have a stronger curvature than the latter part and the particles will drift sideways with spiralling motion.

The region under consideration containing the charged particles is thus subject simultaneously to a magnetic field (directed northward) and to an electric field (directed eastward). Under their joint influence all the charged particles within the region will drift radially outwards (at right angles to both H and F) with the velocity $v = cF/H$. This drift we may call *polarization field drift*. As the positive and the negative space charges are continuously produced at the west and the east sides respectively (by the inhomogeneity of the field) of the region containing the high energy particles, this region will move radially away from the sun, and will form a ray which will be maintained as long as the production continues. The ray or the stream will be laterally polarized in the equatorial plane being traversed by a west-east electric field. It is to be noted that the velocity

in the (nearly) circular paths round the magnetic lines of force, is large in comparison with the velocity of the drift motion.

It may be noted that in the process of polarization the positives and the negatives drift in directions *against* the electric forces acting on them. As such, part of the kinetic energy of the particles (due to nearly circular motions) is transformed into potential energy of the electric field produced by the drifts. The energy depends on the strength of the inhomogeneous magnetic field which produces the polarization and is given by kH (electron-volts) where k is a constant depending on the initial condition. (This omits a relativity correction which is of importance for the high energy particles, e.g. 10^8 eV energy, up to distances of about 10 sun-radii.) For electrons of 10^8 eV energy generated at the surface of the sun, in a field of $H = 30$ gauss, the value of k is $\frac{1}{3}.10^9$ eV/gauss.

It can be shown that the lateral 'inhomogeneity' drift speed of the particles varies as $1/r$, the forward 'polarization field' drift speed as $\sqrt{r}$, the number-density in the stream as $1/r^{\frac{5}{2}}$ and the polarization electric field as $1/r^{\frac{3}{2}}$ (where r is the distance from the centre of the sun). Another characteristic of Alfvén's solar stream, which is of importance in his theory (see Sec. 5), should also be mentioned. In the above we have, for simplicity, considered only such of the high energy particles as are moving in the equatorial plane. If, however, the particles which have velocity components perpendicular to the equatorial plane are taken into consideration, then it is found that though the width of the stream in the equatorial plane is not sensibly increased, that perpendicular to the same is much enlarged. This is because, while the lateral drift does not appreciably broaden the ray, the velocity component perpendicular to the plane causes (in addition to the nearly circular motion with lateral drift) an oscillating motion along the lines of force perpendicular to the equatorial plane. This has the effect of considerably increasing the dimension of the stream perpendicular to the stream.

(*c*) Nuclear processes

Since the interior of the sun must be a seat of nuclear reactions it is very likely that high speed particles produced by such reactions are shot off from the sun. Saha [18] has suggested that nuclear reactions may occur near the sun's surface also. It is known that the solar coronal lines (as shown by Edlen, following a suggestion of Grotrian) are to be attributed to such highly ionized atoms as Fe^{+13} to Fe^{+9}, Ni^{+15} to Ni^{+11} and Ca^{+12} and Ca^{+11}. According to Saha such particles may be produced by nuclear fissions occurring near the surface of the sun. He has also suggested a process by which Fe^{+13} can be produced. It is very probable that nuclear reactions which produce the highly ionized coronal particles, also produce at the same time fast ionized particles.

It is interesting to note in this connection that there is now evidence that cosmic rays (which are very high-energy charged particles) are produced

in the sun during solar flares. In Fig. 1 it will be seen that the cosmic ray intensity record has increased greatly almost simultaneously with the

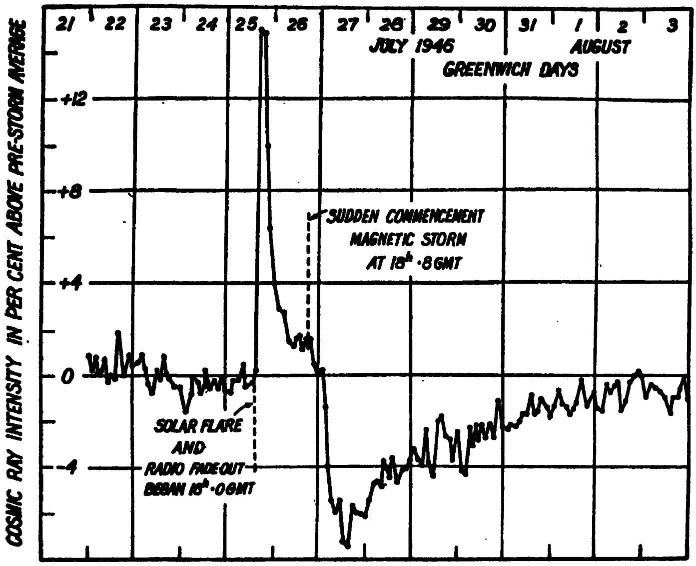


FIG. 1. Illustrating the simultaneity of solar flare and radio fade-out phenomenon on the one hand and the sudden increase of cosmic ray intensity on the other. It will be noticed that the sudden commencement of the associated magnetic storm occurred 26·8 hours later. This indicates that the solar corpuscles responsible for the storm travelled with speed of the order 1000 km./sec. The cosmic ray intensity was recorded at Cheltenham, Maryland, U.S.A.

solar flare, when there was also an increase in ionospheric absorption (radio fade-out). This effect is generally observed in high and middle latitudes but not in the equatorial regions. An interesting point to note is that during the main phase of the magnetic storm the cosmic ray intensity has decreased below the average level. This may be an effect of the magnetic field of the equatorial ring current (see Secs. 3, 4), causing deviation of the cosmic rays from their usual statistical distribution in latitude. It will also be noticed that the associated magnetic disturbance occurred about 24 hours later indicating that the charged corpuscles responsible for such a disturbance travel with speed of the order 1000 km./s. (This figure may be compared with Fig. 78, Plate III in Chapter VI.)

(*d*) Evidence of the presence of corpuscular streams

It has been suggested that a beam of particles, charged or uncharged emanated from the sun ought to be discernible by its dimming effect on the solar disc when the beam is in the line of sight between the earth and the sun. No sign of any such effect has been detected. Two other independent evidences, of the incidence of charged particles during magnetic storms and auroral displays have, however, been obtained.

Solar spectrograms taken at the Mt. Wilson Observatory [19] during the solar cycle ended in 1944, showed that those obtained during magnetic storms indicate two very shallow absorption bands extending about 12 A towards the short wavelength side from about the centre of the *H* and *K* lines. The absorption though small, only about 1 per cent of the background continuum, thus indicates the presence of Ca^+ ions in the interplanetary space speeding towards the earth. (The possibility of detecting such absorptions was first suggested by Chapman [20].) Computations show that the maximum velocities are about 1,000 km./sec. and the mean ones about 600 km./sec. Observations similar to above have also been made at Cambridge, England [19*a*].

A direct proof of the entry of protons into the terrestrial atmosphere has been obtained by the asymmetrical broadening—extension more towards the violet—of the $H\alpha$ (λ 6562) line in the aurora. During a moderate aurora the shift may correspond to a mean velocity of 675 km./s. with a maximum 1,350 km./s. [20*a*]. During intense auroral displays a velocity up to 3,200 km./s. has been recorded [20*b*]. It is to be noted that the asymmetrical broadening is observed only when the direction of observation is along the magnetic lines of force. For observations at right angles to the lines of force there is only a small symmetrical broadening—both towards red and violet—by about 10 A [20*a*]. This may be due to motion of the scattered atoms in the line of sight or to spiral motion round the line of force.

It may also be mentioned that observations show that there is close correlation between incidence of magnetic storms and appearance of bright eruptions on the sun's disc—in which the visible hydrogen and calcium radiations are greatly enhanced. It has further been found that great magnetic storms and auroral displays tend to occur about a day after the bright eruption. It is, therefore, natural to suppose that such storms and displays are caused by charged particles emitted from the sun simultaneously with the bright eruptions.

(*e*) Form of the Solar Corpuscular stream

Irrespective of the process of emission of the charged particles, one can draw certain conclusions regarding the form of the emissions, from considerations of the observed characteristics of the auroras, magnetic storms and associated ionospheric effects which are supposed to be caused by such emissions.

Thus, as first suggested by Maunder [21], the emissions from the solar surface should be in the form of narrow beams. The 27-day recurrence tendency of magnetic storms is then easily explained if the natural assumption is made that the beam emanating radially rotates with the sun and the auroral and magnetic phenomena occur when the beam sweeps over the earth in course of each solar rotation. Chapman [21*a*] has studied the geometry of the beam in detail and draws attention to the fact that owing to rotation of the sun, the beam will be curved in space—the

curvature depending inversely on the velocity of the particle. The effect is thus described by him: 'As the sun rotates the direction of emission changes continuously, and the stream will sweep round like water from a fire-hose whose direction of emission is rotated' (see Fig. 2).

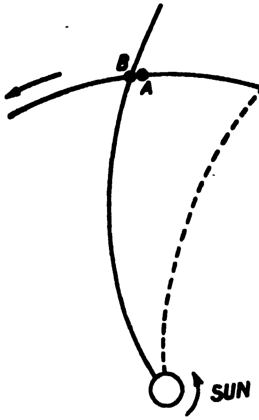


FIG. 2. Illustrating how a stream of particles emitted continuously, radially from a point source on the sun is curved due to the sun's rotation. (The radii of the circles indicating sun and earth are magnified ten-fold in relation to the radius of the earth's orbit.) The speed of the particles is assumed to be 1000 km./sec. The earth moves over the distance *AB* in the time taken by the stream of particles to reach the earth's orbit from the sun. (After Chapman.)

It is also possible to predict some of the characteristic features of the beam from the following considerations. The region of the solar surface emitting the beam must have a synodic period of rotation the same as the recurrence period of magnetic storms [Appendix, Sec. 8]. From observations on sunspots it is found that their synodic rotation periods are 26·9, 27·3 and 28·3 days at latitudes 0°, 15°, and 30° respectively. The beam must therefore be emitted from a region of the solar surface lying not higher than 15° latitude. Again, if it is assumed that the width of the beam is directly connected with the duration of a storm then for a storm lasting 24 hours the width of the beam should be 13·3°; for a 2-hour disturbance it should be 1°. (These calculations are based on the supposition that the emissions travel with the velocity of light. If the emission consists of material particles travelling with velocity less than that of light, the width required would be less.) Finally, for a storm with a sudden commencement the beam must possess an extremely sharp edge.

For reasons discussed below the beam (radial) cannot be a narrow one. The surface rotation of the sun is not constant with time. It varies with the solar activity and increases by about 6 per cent from the epoch of sunspot minimum to that of sunspot maximum. If the magnetic disturbances on the surface of the earth were to be associated with beams of solar radiation emanating from the surface of the sun and participating in its rotation, one would expect the 27-day period of recurrence to show a systematic lengthening and shortening for groups of years of few and many sunspots. No such systematic variation of the 27-day period has yet been detected [22].

We have already referred to the seasonal variation of magnetic activity which shows two maxima, one near the vernal equinox and the other near

the autumnal equinox, when the sun crosses the earth's equatorial plane [Chapter VII, Sec. 7]. Attempt has been made to explain the occurrence of these two equinoxial maxima by considering the tilt of the sun's axis of rotation to the ecliptic [23]. The inclination of the axis with respect to the line joining the earth and the sun is greatest (7·2°) at about September 7 and March 5. At the former we see most of the northern hemisphere of the sun and at the latter most of the southern hemisphere. The beam emissions, which are supposed to start from belts lying between heliographic latitudes 10°–15° within which sunspots occur most frequently, will therefore sweep across the earth mostly in the months of September and March. In the former month the emissions reaching the earth are mostly from spotted belts in the northern hemisphere and in the latter from those in the southern hemisphere. An explanation of the equinoxial maxima may thus be obtained. This view has, however, been subjected to close scrutiny by Chapman and also to rigorous tests by Bartels [24] and is found untenable.

From the above it follows that the beam may be narrow in the equatorial plane, but must be broad in the perpendicular plane so that the frequency of occurrence of magnetic storms would not depend on the heliographic latitude of the earth. It may be recalled that Alfvén's solar stream conforms to this condition (*vide supra*).

Finally, to obviate the Schusterian difficulty, the stream, as suggested by Lindemann [8], must be electrically neutral, consisting of charged particles of opposite signs in equal numbers.

3. CHARGED SOLAR CORPUSCULAR STREAM THEORY (BIRKELAND-STÖRMER)

(*a*) Introduction

As already mentioned in Sec. 1 the idea of Birkeland and Störmer that auroral and magnetic disturbances are caused by streams of charged particles (carrying charges of one sign only) being deflected towards the earth by its magnetic field, labours under the Schusterian difficulty. This difficulty is of a fundamental character and forces one to assume that the stream, as a whole, must be electrically neutral, that is, it must contain charged particles of both the signs in equal numbers. But such a stream will not be deflected by the earth's field, and, as such, will not be able to produce the auroral or the magnetic storm phenomena after the manner envisaged by Birkeland and Störmer. There thus appears to be no other alternative but to abandon the theory. Nevertheless, the analysis of the motion of a single charged particle arriving into the terrestrial magnetic field from a great distance, as was started by Birkeland and later developed in great detail by Störmer, in connection with the development of their theory, is of considerable importance and has many applications irrespective of the correctness of the theory. As such we now proceed to give a brief résumé of this analysis.

(b) Motion of a charged particle in the field of a magnetic dipole

(i) *Störmer's analysis.*—Consider a particle of charge e and mass m moving with a velocity v in a homogeneous magnetic field of intensity H. The particle experiences a force in a direction at right angles to its direction of motion and to H. The magnitude of the force is equal to $Hev \sin \alpha$ where α is the angle between $\vec{H}$ and $\vec{v}$. If the field be uniform and the velocity is at right angles to the direction of the field, then the particle describes a circle of radius ρ given by

$$\frac{mv^2}{\rho} = Hev.$$

The ability of the particle to counteract the deflecting force Hev is directly proportional to its momentum mv and inversely to its charge e. The quantity $mv/e = H\rho = S$ is therefore called the 'stiffness' of the particle. It is equal to the radius of gyration in a magnetic field of unit intensity. Since the force Hev always acts at right angles to $\vec{v}$, the magnitude of the velocity will not in any way be affected even when H changes from point to point.

Let the components of the magnetic field at any point x, y, z be H_x, H_y, H_z. The equations of motion of a negatively charged particle may be written as:

$$m\frac{d^2x}{dt^2} = e\left[H_z\frac{dy}{dt} - H_y\frac{dz}{dt}\right] \quad \ldots \quad \ldots \quad \ldots (1.1)$$

$$m\frac{d^2y}{dt^2} = e\left[H_x\frac{dz}{dt} - H_z\frac{dx}{dt}\right] \quad \ldots \quad \ldots \quad \ldots (1.2)$$

$$m\frac{d^2z}{dt^2} = e\left[H_y\frac{dx}{dt} - H_x\frac{dy}{dt}\right] \quad \ldots \quad \ldots \quad \ldots (1.3)$$

Multiplying these equations by dx/dt, dy/dt and dz/dt respectively and adding, we find that

$$\frac{1}{2}m\frac{d}{dt}\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2\right] = 0$$

or

$$\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2 = \text{constant}.$$

This means, as already stated, that the velocity of the particle is independent of the magnetic field H and remains constant throughout its course. We may therefore introduce the line element ds in place of the time element dt according to the relation $ds = v\,dt$. Eqs. (1) therefore reduce to

$$\frac{d^2x}{ds^2} = \frac{e}{mv}\left[H_z\frac{dy}{ds} - H_y\frac{dz}{ds}\right], \quad \ldots \quad \ldots \quad \ldots (2.1)$$

$$\frac{d^2y}{ds^2} = \frac{e}{mv}\left[H_x\frac{dz}{ds} - H_z\frac{dx}{ds}\right] \quad \ldots \quad \ldots \quad \ldots (2.2)$$

$$\frac{d^2z}{ds^2} = \frac{e}{mv}\left[H_y\frac{dx}{ds} - H_x\frac{dy}{ds}\right] \quad \ldots \quad \ldots \quad \ldots (2.3)$$

Consider now the field H to be that due to the earth. At large distances this field may, with fair approximation, be assumed to be that due to a uniformly magnetized sphere, or to a dipole placed at the origin of the co-ordinate system with its moment M directed along the Z-axis [see Chapter VII, Sec. 1]. For such an elementary dipole the components of the magnetic field are given by

$$H_x = -M\frac{3xz}{r^5} \quad \ldots \quad (3.1)$$

$$H_y = -M\frac{3yz}{r^5} \quad \ldots \quad (3.2)$$

$$H_z = -M\frac{3z^2-r^2}{r^5} \quad \ldots \quad (3.3)$$

where $$r^2 = x^2+y^2+z^2.$$

Substituting the values of H_x, H_y and H_z from Eqs. (3) and putting C for $(Me/mv)^{\frac{1}{2}}$, Eqs. (2) reduce to

$$\frac{d^2x}{ds^2} = \frac{C^2}{r^5}\left[3yz\frac{dz}{ds}-(3z^2-r^2)\frac{dy}{ds}\right] \quad \ldots \quad (4.1)$$

$$\frac{d^2y}{ds^2} = \frac{C^2}{r^5}\left[(3z^2-r^2)\frac{dx}{ds}-3xz\frac{dz}{ds}\right] \quad \ldots \quad (4.2)$$

$$\frac{d^2z}{ds^2} = \frac{C^2}{r^5}\left[3xz\frac{dy}{ds}-3yz\frac{dx}{ds}\right]. \quad \ldots \quad (4.3)$$

If the quantity C which has the dimension of length is chosen as the unit of length, the equations take the form

$$\frac{d^2x}{ds^2} = \frac{3yz}{r^5}\cdot\frac{dz}{ds}-\frac{3z^2-r^2}{r^5}\cdot\frac{dy}{ds} \quad \ldots \quad (5.1)$$

$$\frac{d^2y}{ds^2} = \frac{3z^2-r^2}{r^5}\cdot\frac{dx}{ds}-\frac{3xz}{r^5}\cdot\frac{dz}{ds} \quad \ldots \quad (5.2)$$

$$\frac{d^2z}{ds^2} = \frac{3xz}{r^5}\cdot\frac{dy}{ds}-\frac{3yz}{r^5}\cdot\frac{dx}{ds}. \quad \ldots \quad (5.3)$$

The significance of the unit C is as follows: A particle characterized by m, v and e moving at this distance (C) in the equatorial plane of the dipole, with its velocity directed at right angles to the radius vector joining the particle to the dipole, will describe a circular path round it of radius equal to C. This is easily proved by equating the centrifugal force mv^2/ρ to the electromagnetic force Hev remembering that the intensity of the field at any distance R in the equatorial plane of the dipole is given by M/R^3.

Using cylindrical co-ordinates R, ϕ, z as indicated in Fig. 3, we may write

$$x = R\cos\phi,\ y = R\sin\phi \text{ and } z = z.$$

By multiplying Eq. (5.1) by sin ϕ and (5.2) by cos ϕ and subtracting we get

$$\frac{d}{ds}\left(R^2\frac{d\phi}{ds}\right)=\frac{d}{ds}\left(\frac{R^2}{r^3}\right). \quad .. \quad .. \quad .. \quad (6)$$

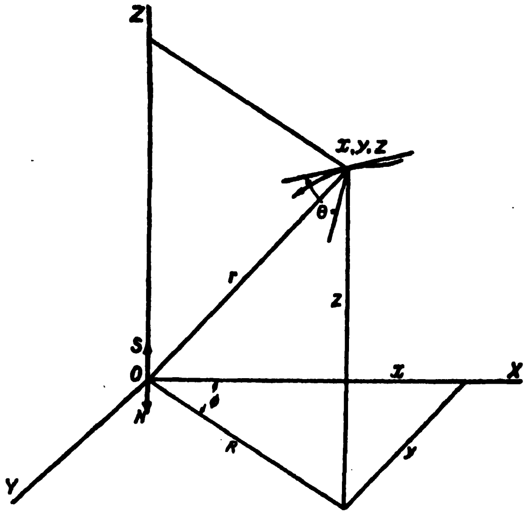


FIG. 3. System of co-ordinates.

This, on integration, gives

$$R^2\frac{d\phi}{ds}=\frac{R^2}{r^3}+2\gamma, \quad .. \quad .. \quad .. \quad .. \quad (7)$$

where γ is a constant of integration which can have any value from $-\infty$ to $+\infty$. Similarly multiplying Eq. (5.1) by cos ϕ and Eq. (5.2) by sin ϕ and adding we get

$$\frac{d^2R}{ds^2}=\frac{1}{2}\frac{\partial Q}{\partial R}, \quad .. \quad .. \quad .. \quad .. \quad (8)$$

where

$$Q=1-\left(\frac{R}{r^3}+\frac{2\gamma}{R}\right)^2. \quad .. \quad .. \quad .. \quad (9)$$

Also from Eq. (5.3),

$$\frac{d^2z}{ds^2}=\frac{1}{2}\frac{\partial Q}{\partial z}, \quad .. \quad .. \quad .. \quad .. \quad (10)$$

and therefore

$$\left(\frac{dR}{ds}\right)^2+\left(\frac{dz}{ds}\right)^2=Q. \quad .. \quad .. \quad .. \quad (11)$$

Solving Eqs. (7), (8) and (10) for R, ϕ and z and remembering Eq. (9), we get the trajectory of the particle.

Since the constant γ figures prominently in the discussion that is to follow it would be useful to understand its physical significance. If the quantity R^2/r^3 (which appears on the right hand side of Eq. (7)) be put equal to a constant α, then the resulting equation is that of a line of force of the

dipole. Further, since $R\,d\phi/ds$ is the component velocity along the tangent to the circle drawn round the Z-axis through the instantaneous position of the particle (remembering that ds can be regarded as an element of time), the quantity $R(R\,d\phi/ds)$ on the left hand side of Eq. (7) is the component of the moment of momentum of the particle along the Z-axis. As the particle moves it cuts the lines of force and, along its path, the constant α defining the lines of force, as also the moment of momentum above defined, change. Eq. (6) therefore states that along the trajectory, the rate of change of moment of momentum (along the Z-axis) is equal to the rate of change of the constant α defining the lines of force. At a great distance, i.e., where r is many times the unit of length C, R^2/r^3 is negligible and the meaning of the constant γ is then simplified. It is then the component of the moment of momentum of the particle along the dipole axis and is a constant of motion.

Inspection of Eq. (7) and of Eqs. (8) and (10) shows that the motion of the particle in space may be regarded as consisting of two parts. Imagine the ROZ plane (Fig. 3) in which the particle is situated to swing round the Z-axis in such a way that the particle is always in this plane. The motion then is resolvable into:

1. Motion in the plane ROZ with R and Z as functions of time (i.e., of s) according to Eqs. (8) and (10).
2. Motion of the plane itself with ϕ as function of time (i.e., of s) according to Eq. (7).

Trajectories in space can thus be determined by first integrating Eqs. (8) and (10) and then finding ϕ from Eq. (7). Rigorous integration which is necessary for a quantitative investigation of the trajectories is, however, difficult and involves extremely laborious computations. Störmer has evolved certain relatively simple methods of graphical integration which can be employed with advantage when a not very exact view of the trajectories is needed. The more laborious task of numerical integration is resorted to when the greatest accuracy is aimed at. Mechanical integration has also been adopted with the help of the 'Bush Differential Analyser' and many thousands of trajectories calculated. But even without actual integration the motion of the particle can be understood and several important conclusions drawn from the above equations.

Motion in the meridian plane ROZ.—According to Eqs. (8) and (10), motion in the rotating plane ROZ is derivable from the function $\frac{1}{2}Q$, s being considered as time. It is best understood with reference to the so-called level lines of Q, the space derivative of which gives the force acting on the particle. The force is thus at right angles to the curves for which $Q =$ constant, and acts in a direction in which Q is increasing. The constant-Q curves can be easily drawn with the help of Eq. (9) for any given value of the integration constant γ (see Figs. 4–6). Thus

$$\sqrt{1-Q} = \frac{2\gamma}{R} + \frac{R}{r^3} = k, \text{ a constant.} \quad \ldots \quad \ldots \quad (9.1)$$

If ψ be the angle between the Z-axis and r, we have $R = r \sin \psi$ and

$$k = \frac{2\gamma}{r \sin \psi} + \frac{\sin \psi}{r^2},$$

whence

$$r = \frac{\gamma \pm \sqrt{\gamma^2 + k \sin^3 \psi}}{k \sin \psi} \qquad \text{.. .. ..} \qquad (12)$$

For a given value of γ the various level lines for which $Q =$ constant, can be drawn by assigning different values to k.

Motion of the plane ROZ.—The motion of the plane ROZ is derivable from Eq. (7) which may be written in the form

$$\frac{d\phi}{ds} = \frac{2\gamma}{R^2} + \frac{1}{r^3}.$$

It should be noted that the motion of the plane may not always be in the same sense; it may oscillate to and fro. If γ is > 0 the quantity on the right hand side will be positive and ϕ will always increase with time. If however γ is negative, equal to $-\gamma_1$ say, the right hand expression can assume zero or negative values. It is easy to see that ϕ will be increasing up to the limit given by

$$\frac{2\gamma}{R^2} + \frac{1}{r^3} = 0,$$

which is the equation of the curve for $Q = 1$, i.e., $k = 0$, as is evident from Eq. (9.1). Thus ϕ will be increasing as long as the particle is within the toroidal surface generated by

$$r = \frac{\sin^2 \psi}{2\gamma_1}.$$

When the particle moves out of this space, ϕ will decrease, that is, the plane will be rotating in the opposite sense. One may, with a little imagination, picture how the trajectory of the particle moving in the plane ROZ will behave when the plane itself changes its sense of rotation.

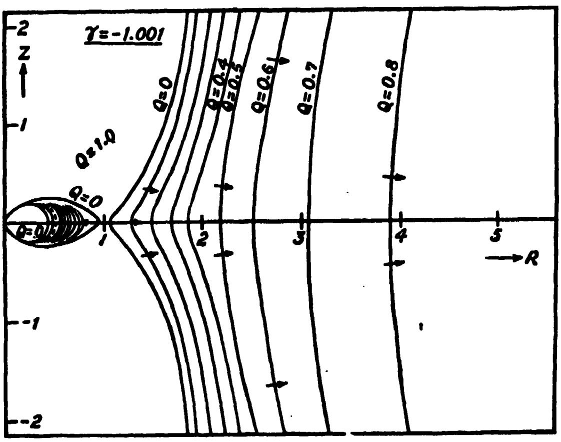


FIG. 4. Constant-Q curves for $\gamma = -1{\cdot}001$. (After Störmer.)

Allowed and forbidden spaces.—In Figs. 4–6 the constant-Q curves are drawn for three different values of γ. It will be noticed that the lines $Q = 0$, form boundaries, as it were, on the one side of which the level lines are confined. In other words the trajectories (from infinity) are also con-

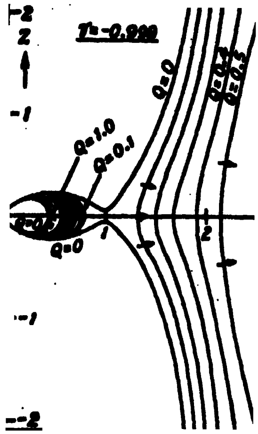


FIG 5. Constant-Q curves for $\gamma = -0{\cdot}999$. (After Störmer.)

fined in a space of which the boundary is given by $Q = 0$. This may also be seen from the following consideration. We have

$$Q = 1 - \left(\frac{R}{r^3} + \frac{2\gamma}{R}\right)^2. \qquad \ldots \quad \ldots \quad \ldots \quad (9)$$

If θ be the angle which the direction of motion of the particle (velocity v)

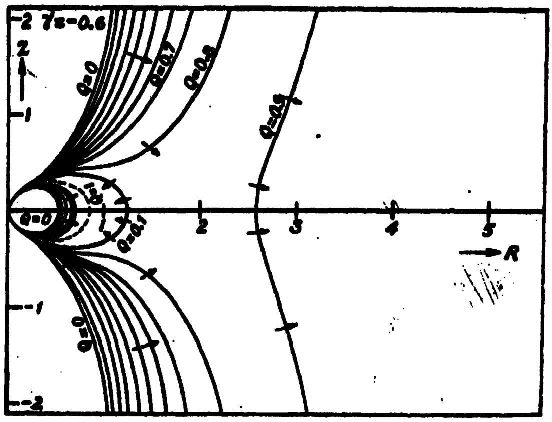


FIG. 6. Constant-Q curves for $\gamma = -0{\cdot}6$. (After Störmer.)

at any instant makes with the rotating plane ROZ, then, remembering that $ds = v\,dt$,

$$\sin\theta = R\frac{d\phi}{ds}.$$

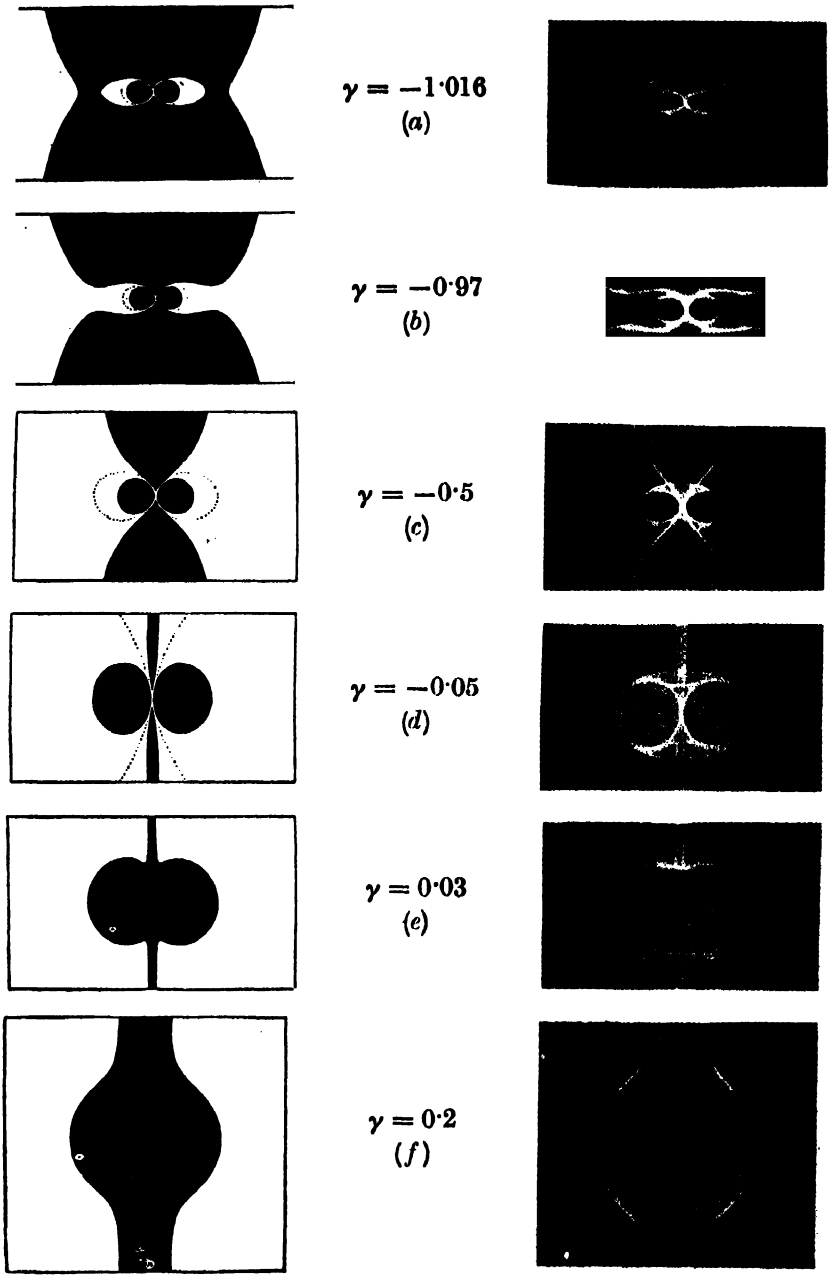


FIG. 7. Illustrating the Q_γ spaces for different values of γ in which the trajectories of a charged particle moving in the magnetic field of a dipole are confined. On the left, the dark regions are meridional sections of the forbidden spaces; the white regions represent the allowed spaces in which the trajectories are confined. On the right, these spaces are pictured in three dimensions. The whitish spaces are forbidden and the dark ones are allowed. (After Störmer.)

Eq. (7) therefore becomes

$$R \sin \theta = \frac{R^2}{r^3} + 2\gamma;$$

or

$$\sin \theta = \frac{R}{r^3} + \frac{2\gamma}{R} = k.$$

Thus we have, from Eq. (9),

$$Q = 1-(\sin \theta)^2$$

and, since the maximum and minimum values of $\sin^2 \theta$ are 1 and 0 respectively, Q must lie between 0 and 1; that is, the traces of the trajectories on the rotating plane are confined in the region

$$1 > Q > 0.$$

The whole space round the dipole can thus be divided into allowed and forbidden spaces, the boundary separating the two depending on the value of γ chosen. The meridional sections of the boundary (highest level) are given by the curves obtained by putting $Q = 0$, i.e., $k = \pm 1$ in Eq. (12). The bounding surfaces, i.e., the surfaces within which the trajectories are confined in space are generated by the revolution of these curves round the Z-axis. A few of these surfaces separating the 'forbidden' from the 'allowed' space are shown in Fig. 7 (Plate I) for six different values of γ ranging from $\gamma = 0{\cdot}2$ to $-1{\cdot}016$. In the figures on the left the dark regions show in section the 'forbidden' spaces while the white regions represent the allowed spaces in which the trajectories are confined. On the right these spaces are pictured in three dimensions as formed by the revolution of the figures on the left round the dipole axis. Here the whitish spaces are forbidden and the dark ones are allowed.

Störmer calls the spaces in which the trajectories are confined Q_γ regions and has investigated them in great detail. He has shown that these regions may be classified into three distinct types according as the value of the integration constant γ falls in any of the following ranges

(i) $\gamma < -1$

(ii) $-1 < \gamma < 0$

(iii) $0 < \gamma$.

The following characteristics of the Q_γ regions for values of γ lying within these ranges may be noted in Fig. 7 (Plate I).

(i) $\gamma < -1$. For this case the allowed Q_γ space consists of two parts one separated from the other; the direct path from infinity to the dipole is blocked.

(ii) $-1 < \gamma < 0$. For this case the two allowed Q_γ spaces unite, i.e., the origin is connected to infinity by free Q_γ space. Trajectories from infinity can reach the origin.

(iii) $0 < \gamma$. For this case the Q_γ space from infinity is blocked at the centre of the dipole. Trajectories from infinity cannot reach the origin.

The trajectories of charged particles moving in the field of a dipole, are of infinite variety and of extremely complicated shapes. They may, however, be conveniently classified into the following four types:

(1) Orbits reaching the earth from infinity, as also orbits proceeding from the earth to infinity.

(2) Orbits from infinity unable to reach the earth and receding away from it after approaching a minimum distance; also orbits starting from the earth and falling back on it being unable to proceed to infinity.

(3) Periodic orbits in which, as the name implies, the charged particles may move round the dipole indefinitely.

(4) Asymptotic orbits which approach the periodic orbits asymptotically.

(ii) *Störmer's wire models of trajectories; curve of precipitation.*—In order to obtain a realistic view of the orbits in space of corpuscles travelling towards the dipole, Störmer has constructed many wire models of the trajectories. Such a model is shown in Fig. 8. It will be noticed that by far the largest number of trajectories turn away from the earth on nearing it. These, we shall see later, contribute to what Störmer calls a 'ring current' round the earth. Fig. 9 is the picture of another wire model illustrating a few trajectories which, emanating from a point source (in the sun) reach the earth. It also shows how such orbits strike the earth only at certain definite spots near the poles. This model also serves to explain, according to the Birkeland-Störmer hypothesis, certain characteristic features of the auroras. The trajectories, for instance, strike on the night side of the earth which is in conformity with the observed occurrence of auroral maxima during night. Again, some trajectories pass right round the earth before striking it. These contribute partly to the ring current referred to above. We may also refer to another wire model

FIG. 8. Störmer's wire model of trajectories. Note that by far the largest number of the trajectories turn away from the earth on approaching it.

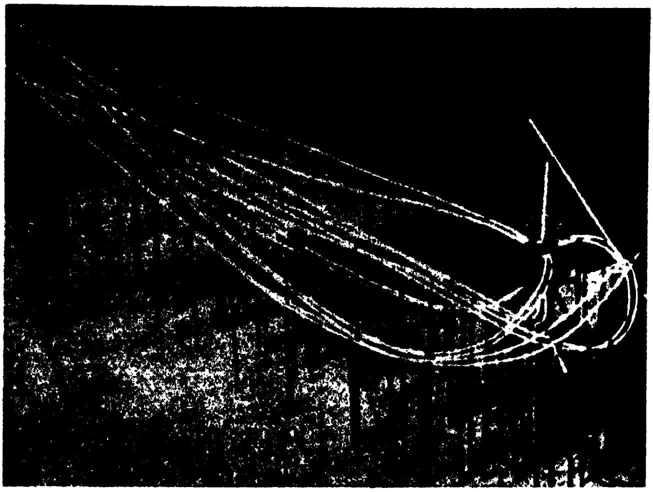


FIG. 9. Störmer's wire model of trajectories emanating from a point source which are able to reach the earth. Note that some of the trajectories strike on the night side of the earth.

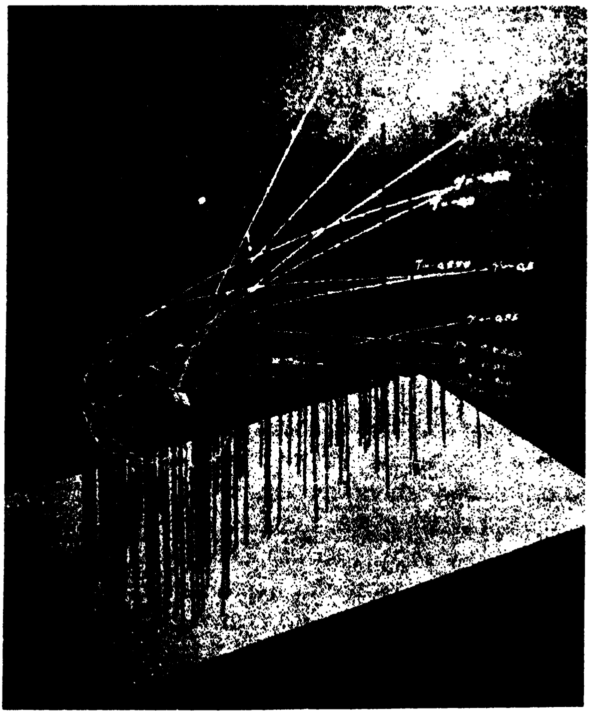


FIG. 10. Störmer's wire model illustrating trajectories (each with a different value of γ) coming from different directions and striking the earth. The trajectories all start parallel to the same meridional plane.

(Fig. 10) in which the trajectories (each with a different value of γ) come from different directions corresponding to different positions of the sun with respect to the axis of the earth magnet. They all, however, start parallel to the same meridional plane, i.e., plane through the magnetic axis. The distribution of the points at which the trajectories cut the surface of a small sphere round the dipole is shown in Fig. 11. It will be seen that the points of intersection are arranged in the shape of a spiral

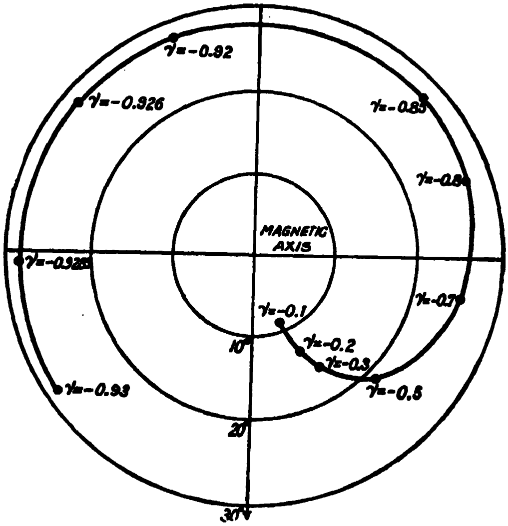


FIG. 11. Curve of precipitation showing the distribution of the points at which the trajectories of Fig. 10 cut the surface of a small sphere round the dipole. (After Störmer.)

round the pole called the curve of precipitation. It will further be noticed that as the spiral proceeds outward, a small change in the γ value (which amounts to a small change in the position of the emitting point) produces a large change in the position of the point of incidence.

(iii) *Experiments with 'Terrella'.*—Apart from the applicability or otherwise of Störmer's theory to explain auroral phenomena, its correctness, so far as the mathematical analysis is concerned, has been subjected to experimental tests. Birkeland [1] who was the originator of the theory tried to reproduce in the laboratory the condition as would exist when a solar beam of charged particles approaches the earth magnet. He constructed a large vacuum-chamber and placed in its centre an iron sphere surrounded by a magnetizing coil (Fig. 12). On exposing the sphere to cathode rays produced in the chamber, it was found that the cathode rays would strike the sphere only when a certain degree of magnetization and a certain velocity of the cathode rays were attained. In some experiments the cathode rays precipitated on the sphere—or the 'Terrella' as it was called by Birkeland, in patches and in others in the form of spirals round the magnetic poles. Birkeland further demonstrated that in some cases the cathode ray bundle

would go round the sphere and strike the Terrella on the night side. In Fig. 12 the concentration of the cathode rays round the magnetic pole is

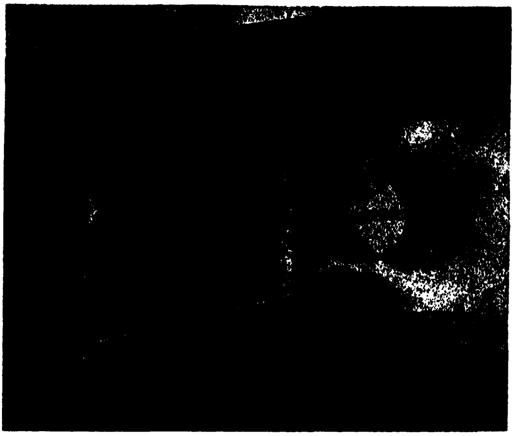


FIG. 12. Birkeland's experiments with 'Terrella'. A magnetized sphere in a vacuum chamber is exposed to cathode rays. Note the dark toroidal space round the sphere within which the cathode rays cannot enter.

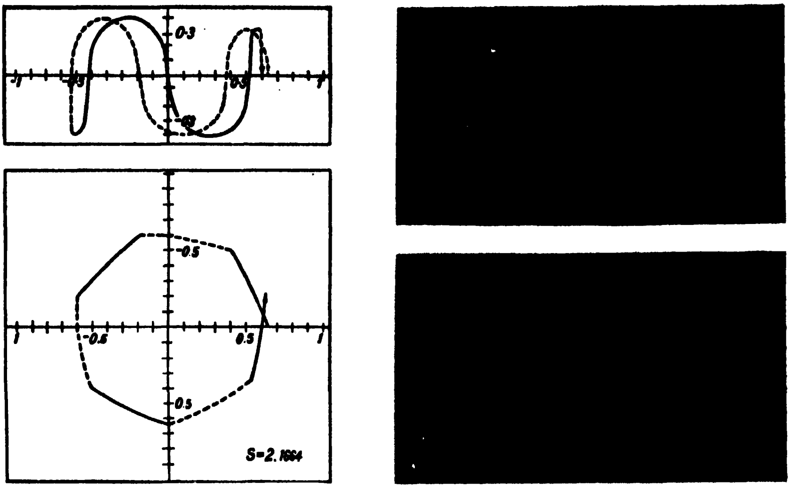
clearly visible, as also the dark toroidal space surrounding the sphere which is forbidden for the cathode rays.

In Birkeland's experiments the cathode ray bundle was fairly wide and the emitting surface was also large. Brüche [25] repeated the experiments using thin narrow bundles of rays—'thread rays'. The electrons had energy of 200 volts (8,400 km./sec.) and the pressure in the discharge tube 0·001 mm. of Hg. Under such condition the positive ions formed along the path of the rays served, by electrostatic action, to check the dispersion of the electrons so that the rays maintained a small cross-section for length up to about 1 metre. The rays were self-luminous and were visible along their entire length of path. The most remarkable experiments of Brüche [26] are those in which some of the periodic orbits predicted by Störmer's theory were reproduced (Fig. 13).

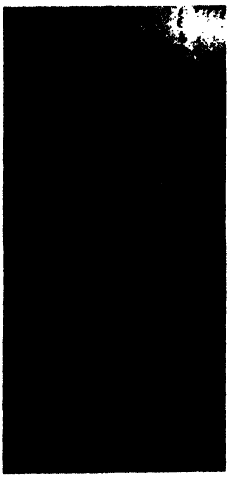
Another important experimental verification is that regarding the effect of Störmer's 'ring current' in the magnetic equatorial plane. Theoretically, such a 'ring current' would cause the zone of precipitation of the charged particles to move towards lower latitudes. Experimental verification of this is shown in Fig. 14. In (*a*) of this figure there is no current in the ring while in (*b*) it is energized by a current such that the direction of its field is opposed to that of the 'Terrella'. It will be noticed that in the latter case the zone of precipitation on the Terrella has shifted slightly equatorwards.

Experiments similar to those of Birkeland and of Brüche have also been performed by Malmfors [27] in connection with Alfvéns theory of magnetic storms and auroras (see Sec. 5).

(iv) *Applications to auroral and magnetic storm phenomena.*—The main results of Störmer's work—so far as their applications to auroral phenomena



(*a*)

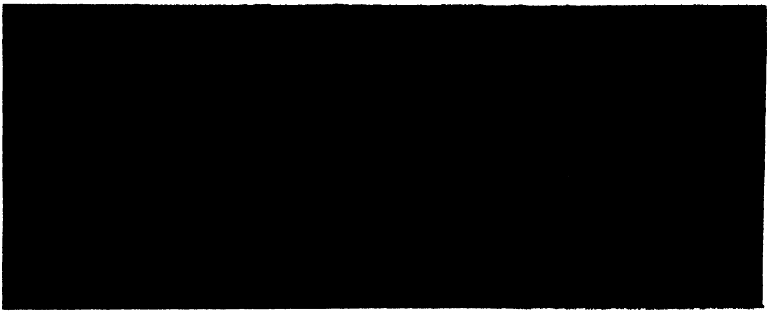


(*b*)

FIG. 13. Brüche's experimental verification of periodic trajectories—(*a*) Oscillating path of the thread rays. *Left*—Path calculated by Störmer. *Right*—Path experimentally observed by Brüche. (*b*) Thread rays as used by Brüche.

and magnetic storms are concerned—may thus be summarized: Of the charged particles ejected from the sun by far the largest number is unable to reach the earth being deflected away by its magnetic field (Fig. 8). Those which reach the earth, do so only round the polar regions (Figs. 9 and 10). These produce auroral displays and the polar magnetic storms by bombardment of the terrestrial atmosphere. The trajectories of the deflected particles are crowded round the magnetic equatorial plane at a great distance from the earth, comparable to the moon's orbit. The motions of the charged particles in these trajectories constitute a giant ring

current round the earth and cause the world-wide magnetic storms. This ring current also exerts a deflecting influence on the particles which reach the earth, causing a small shift of the auroral belt towards lower latitudes.



(a) (b)

FIG. 14. Illustrating the effect of an equatorial ring current on the zone of precipitation of charged particles on a magnetized sphere. (a) Ring is not energized. (b) Ring is energized by current. Note that in the latter case the zone of precipitation has shifted slightly equatorwards. (After Brüche.)

However, the fundamental objection to the Birkeland-Störmer theory, as first pointed out by Schuster, always remains. In the theory the stream consists of charged corpuscles of one sign only. But to produce the observed geomagnetic disturbances and auroras the number density of charged particles in the stream has to be so high that mutual electrostatic repulsion would disperse the particles in the stream long before they reach the earth.

There is also another difficulty which, though not so fundamental as the above, is nevertheless quite serious. This is regarding the angular radius of the auroral zone maximum. From the analysis given above it easily follows that the charged particles coming from outside are mainly concentrated near the poles. The maximum equatorial extension of the point of precipitation is given by $\psi_0 = \sqrt{2a/C}$. (See Chapter I, Sec. 5e). ψ_0, therefore, is a function of the momentum of the particles. As ψ_0 is known to be about 25°, the momentum necessary for precipitation at this angular distance can be calculated. It is, however, found that for the masses of the likely particles (electron, H^+, Ca^+), the required speed is either too large (electron, H^+) or too small (Ca^+). That is, either the particles are so penetrating (electron, H^+) that they are not able to ionize the atmosphere at the required level, or they are so slow (Ca^+) that they are unable to reach down to this level (viz. round 100 km.). To avoid this difficulty Störmer, as mentioned above, postulated the existence of a 'ring-current' in the equatorial plane at distance of several earth-radii. The effect of the magnetic field of this current was to produce equatorwards deflection of the precipitating particles (Fig. 14). The existence of a ring-current, associated with magnetic storms, is also now generally recognized. However, the intensity of the magnetic field of such current is too small to produce the

effect as supposed by Störmer. We shall see in Sec. 4 how the angular radius of the auroral zone is explained by the neutral solar corpuscular stream theory.

4. NEUTRAL SOLAR CORPUSCULAR STREAM THEORY (CHAPMAN-FERRARO-MARTYN)

(*a*) Introduction

As already mentioned, to obviate the Schusterian difficulty Lindemann [8] suggested that the solar stream may be electrically neutral consisting of charged corpuscles of opposite signs in equal numbers. A new difficulty, however, appears with this assumption. A neutral stream would, as a whole, suffer too little deflection in the earth's magnetic field. It then becomes difficult to explain how the charged particles would concentrate in the regions round the magnetic poles of the earth to produce the auroral and magnetic disturbance phenomena. However, closer scrutiny shows that when the neutral beam invades the region of the earth's magnetic field, the paths of the positive and the negative particles may become differentiated through the action of the field. The Schusterian difficulty is then not crucial. In what follows we will describe two theories, one due to Chapman and Ferraro and as later extended by Martyn, and the other to Alfvén based on the assumption that the sun occasionally emits streams of charged particles of both signs so that the resulting beam is electrically neutral. It appears that the Chapman-Ferraro-Martyn theory will be the surviving theory.

(*b*) Chapman and Ferraro's work

From various considerations the approximate nature and properties of the neutral stream may be predicted. The velocity with which the particles stream out of the sun is of the order of 10^8 cm./sec. At a distance from the sun equal to the radius of the earth's orbit, the density of the stream will lie between 2×10^9 and 20 ions per cm.^{3} The velocity (relative to the earth) of the lateral surface of the stream along the earth's orbit will be about 0·056 earth-radius per sec.

If solar rotation be taken into account, then the beam will be found to have a curved form though the particles are emitted nearly radially from the active spot of the sun (see Fig. 2). The width of the stream will of course depend on the area of the spot but an approximate estimate may be made from the following considerations. It may be assumed that the magnetic field of the earth begins to be affected by the stream from a distance (perpendicular to the earth's magnetic axis) where the field has a value $0{\cdot}3\gamma$, and that the time during which the sphere of magnetic influence remains enveloped by the stream is a measure of the duration of the magnetic storm. The diameter of this sphere is about 100 earth-radii. If

the average duration of a storm is taken as a day, the diameter of the stream comes out to be about 5,000 earth-radii.

The effect of the terrestrial magnetic field on the neutral stream, as it laterally approaches the earth, may now be considered. For the sake of simplicity the stream may be assumed cylindrical and of circular section. The magnetic field will tend to deflect the ions and electrons in the stream differently and to separate them. But the electric field set up by this separation will resist the tendency and the net result will only be a slight polarization of the stream. The means that there will be a surface charge-distribution and, in the case of the cylindrical stream under consideration, an electric field will be set up outside the stream in addition to the one inside. The external electric field will cause charges of either sign to escape from the surface of the stream, the leakage being immediately made good from the body of the stream. However, the density of any such dispersing cloud of charges will be too small to exert any appreciable magnetic field. Further, though many of the escaped electrons will find their way to the polar regions guided by the earth's magnetic lines of force, they will not be able (having velocities of the order 1,000 km./s. only) to penetrate sufficiently deep into the atmosphere. Hence, the polarized neutral beam, with induced surface charges escaping into space, will not produce any appreciable auroral or geomagnetic effects. (It will, however, be seen in the next section that these phenomena, which are neglected as only of secondary importance, are precisely those on which Martyn bases his explanations of the auroral and geomagnetic effects by the Chapman-Ferraro theory.)

But, as the stream advances into regions of gradually increasing magnetic intensity, other phenomena begin to appear which have direct bearing on the production of magnetic storms. Consider first the beam as rigid. Since the radius of the stream is assumed to be about 5,000 earth-radii, the surface of the stream, with reference to the earth, will appear as plane when it is a few tens of earth-radii from the earth. The approach of the beam may, therefore, be considered as the approach of a conducting plane (only slightly inclined to the plane containing the sun's axis and the earth) into the field of the terrestrial magnetic dipole. The conductor moving in the variable field will cause electric currents to be produced within a thin surface layer, the effect of which will be to oppose the interpenetration of the conductor by the external magnetic field.

The magnetic effects of the induced current system may be represented, as those due to an image-dipole situated inside the stream, at the same perpendicular distance from the surface as the earth-dipole is outside. The effect of this image-dipole (i.e. of the field of the induced current system) would be to compress the lines of force of the earth-dipole on the side near the stream (Fig. 15). The result would be an increase in the magnetic intensity round the earth which is characteristic of the first phase of the magnetic storm.

It is possible to calculate without much difficulty the change of energy in the system due to the approach of the magnetic doublet near the plane

conductor. The magnetic field of the induced currents in the surface layer of the conductor will be the same as that due to a mirror image of the

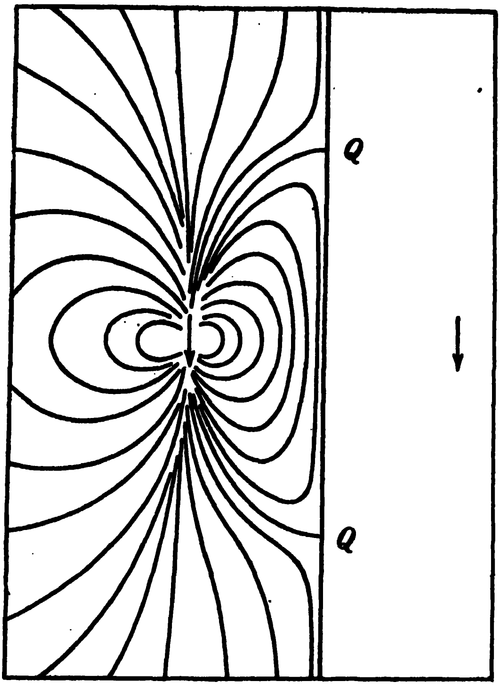


FIG. 15. Illustrating the magnetic field due to the combined effect of the earth-dipole (arrow on the left) and the electric current system induced on the advancing plane conducting surface QQ of the neutral ionized stream. The magnetic effect of the latter (on the left of QQ) is the same as that due to the image doublet as shown by the arrow on the right. (After Chapman.)

terrestrial magnetic doublet produced inside the conductor. Thus the mechanical force between the conductor and the doublet is the same as that between the real doublet and the image doublet and is given by

$$F = \frac{3M^2}{(2r)^4}$$

where $2r$ is the distance between the two doublets and M the magnetic moment of each. The work done in bringing the image doublet from infinity to r is

$$\int_{\infty}^{r} F\,dr = M^2/16r^3.$$

At a distance of ν earth-radii ($r = \nu a$) the work done, or the energy stored, is equal to

$$\frac{M^2}{16\nu^3 a^3} = 1{\cdot}7 \times 10^{24}/\nu^3 \text{ ergs}$$

since a, the radius of the earth $= 6{\cdot}37 \times 10^8$ cms. and $M \doteqdot 8{\cdot}5 \times 10^{25}$ C.G.S. e.m.u.

Thus when the stream surface is at a distance of about 4 to 5 earth radii the increase of energy is the same as that of an average storm (10^{22} ergs). The calculation is of course very approximate because the stream

has been taken as a conducting rigid body approaching the earth without any distortion, whereas, as will be seen presently, the surface of the stream is profoundly modified by the action of the field of the terrestrial magnetic doublet. The treatment, however, indicates how the approach of the conducting stream may be responsible for the increase of the horizontal magnetic intensity during the first phase of the storm.

In order to explain the second or the main phase of magnetic storms the mechanical force acting between the induced current in the surface of the stream, which is no longer supposed to be rigid, and the magnetic field of the earth has to be considered. This force will tend to retard the motion towards the earth of the current bearing surface. The retarding force per unit surface area of the current layer is obviously proportional to the product of the tangential magnetic intensity, H_S and the current density, i. Since the latter in its turn is approximately proportional to H_S and to the normal velocity v_N of the surface, the retarding force is proportional to $v_N \times H_S^2$. Or, since H_S varies as r^{-3} it is proportional to $v_N \times r^{-6}$. A little consideration will show that this retarding force in the immediate neighbourhood of the earth will distort the stream surface and produce a hollow the equatorial section of which will be roughly parabolic as shown in Fig. 16.

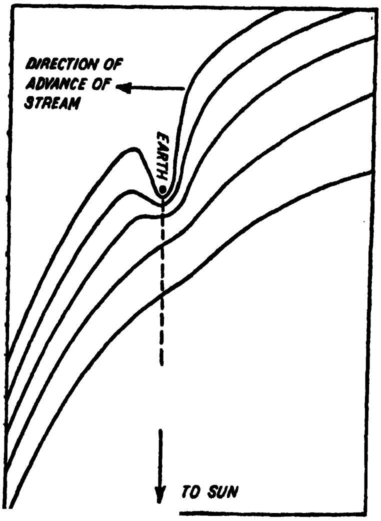


FIG. 16. Illustrating the formation of a hollow space in an advancing corpuscular stream when nearing the earth magnet. The hollow is roughly of parabolic form. (After Chapman and Ferraro.)

As the stream advances into the magnetic field of the earth and the hollow develops, the positive ions and the electrons in the surface layer also advance with the stream and are deflected in opposite directions by the magnetic field of the earth. The electrons and ions thus separated will produce positively and negatively charged layers on the walls of the hollow as shown in Fig. 17. The positively charged layer will be opposite the morning side of the earth (BB′) and the negatively charged one on the evening side (CC′). Due to the electric field thus produced the charges will tend to leap the gap between the walls of the hollow. The current circuit can, however, be completed only at a distance where the radius of gyration of the charged particle (mv/He) is comparable to the breadth of the gap. Since at any distance the radius of gyration of ions is much greater than that of the electrons the current flow will consist mainly of ions. The electrons will drift away under the influence of the crossed electric and magnetic fields.

A feeble ionic current leaping the gap at a great distance will build up stronger currents nearer the earth. This is because the magnetic field produced by the current is in a direction opposite to the terrestrial magnetic field on the side nearer the earth. The terrestrial magnetic field is thus weakened and as a consequence the radius of gyration of the ions increases and facilitates the growth of the current across the gap. The current flowing across the gap may be completed through the main stream or may close upon itself and form an isolated current ring round the earth. In any case this ring current will produce world-wide depression of H, the main phase of the magnetic storms.

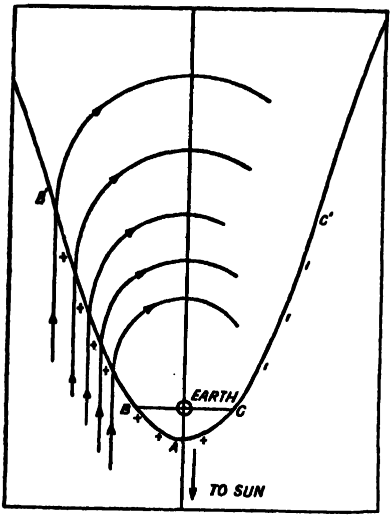


FIG. 17. Illustrating a later development of the hollow shown in Fig. 16. Charges of opposite signs develop on the wall of the hollow and tend to leap the gap along the curved paths. (After Chapman and Ferraro.)

In the above attempt has been made to explain only the initial and the main phase of the magnetic storms. In regard to the other features of magnetic disturbances and production of the auroral phenomena, Chapman [28] in a later review of the theory remarks that it may be supposed that some of the surface charges that leave the hollow (but *do not* bridge the gap and contribute to the ring current) may find their way to the earth's atmosphere in high latitudes, perhaps with increased speed. During their passage to the earth the charged particles will be subject both to electrostatic and electromagnetic forces and these may jointly determine the position of the auroral zones as also the complicated features of auroral precipitation.

(*c*) Extension of the Chapman-Ferraro theory (Martyn)

(i) *Introduction.*—The Chapman-Ferraro theory, as outlined above, has been extended by Martyn [12] in an attempt to explain the morphologies of magnetic storms, auroras, and the associated ionospheric variations. Martyn has made use of hydrodynamical analogies which suggest strongly that the neutral stream will *completely* enclose a hollow space surrounding the earth. (In this respect he departs from the Chapman-Ferraro theory, in which there is a space extending to infinity behind the earth unfilled by the stream). He considers the problem as akin to that of the steady state of motion of a fluid around a submerged obstacle, the pressure gradients in which are replaced by the body forces due to the interaction between electric currents in the stream and the earth's magnetic field. According to these ideas the ring current round the hollow is simply a manifestation

of the forces which constrain the stream to flow in a curved course round the walls of the hollow ('forbidden') space surrounding the earth.

Starting with these ideas and making simple, plausible assumptions Martyn first deduces the radius of the 'forbidden' space (about 5·5 earth radii) and then shows that the lines of force of the earth-dipole which touch this space in the equatorial plane will cut the earth's surface at an angular distance of about 25° from the poles. This result is significant, as this also is approximately the angular radius of the auroral zone: Charged particles ejected from the region of the ring current will bombard the earth's atmosphere in the auroral zone. Martyn also shows that these particles have sufficient energy to penetrate to auroral levels. A brief account of Martyn's work is now given.

(ii) *Estimation of the size of the hollow and the number density of the particles in the solar stream.*—This can be made from a consideration of the magnitude of the rise in H during the initial phase of a magnetic storm, when the neutral stream is invading the region of the earth's magnetic field. As already explained, the magnetic effect of the current system induced on the conducting surface (nearly plane for the portions near the earth) of the stream may be represented as that due to an image-dipole inside the stream (see Fig. 15). If z be the distance of the earth-dipole from the surface of the stream, then the distance between the earth-dipole and its image is $2z$. Hence, taking the average rise in H at the surface of the earth to be 25γ above its mean value (0·33 gauss) we have

$$0{\cdot}33/(2z)^3 = 25\times 10^{-5}, \text{ or, } z = 5{\cdot}5 \text{ earth radii.}$$

To estimate the number-density N of the particles (moving with velocity u) we first make another estimate of the dimension of the hollow which involves N. We make the justifiable assumption that the pressure exerted by the solar stream (i.e. the energy density of the stream $\frac{1}{2}Nmu^2$) is equal to the pressure of the lines of magnetic force at the surface of the hollow (i.e. $H^2/8\pi$). Since the intensity of the terrestrial field at distance z is $0{\cdot}33/z^3$, we have

$$\tfrac{1}{2}Nmu^2 = \frac{(0{\cdot}33)^2}{8\pi z^6}, \quad \text{or,} \quad z = \left(\frac{0{\cdot}2}{8\pi Nmu^2}\right)^{\frac{1}{6}}. \quad \ldots \quad \ldots \quad (1)$$

The size of the hollow is thus seen to vary little with the number-density N or the mass of the particles comprising the solar stream. If u is taken as 1,000 km./s. and m that of the hydrogen atom, then

$$z = 8{\cdot}8\left(\frac{1}{N}\right)^{\frac{1}{6}}.$$

If we put $z = 5{\cdot}5$ earth-radii as deduced above we get the value of N to be 20.

(iii) *Angular radius of the auroral zone.*—Consider the lines of force of the earth-dipole which touch the hollow in the equatorial plane. The

angular distance (θ_1) from the pole (co-latitude) of the points where they cut the surface of the earth is given by

$$\frac{1}{\sin^2\theta_1} = 5{\cdot}5.$$

Thus $\theta_1 = 25°$. This is nearly the angular radius of the auroral zone. The bearing of this will be discussed presently.

(iv) *The ring-current and its cross-section.*—The flow of the solar stream, on the lateral inside surface of the hollow near the equatorial plane, is in a roughly circular path of radius r ($= 5{\cdot}5$ earth-radii). According to Chapman and Ferraro a radial polarization field E (directed inside the stream from the surface) is necessary for the stability of the flow in this curved path. If u_i be the velocity (taken positive when westward) of the ions then the resultant force due to the polarization field E and the magnetic field H is $eE-eu_iH$. Equating this with the centrifugal force we have

$$-m_i u_i^2/r = eE-eu_iH.$$

Similarly, if u_e be the velocity of the electrons (again taken as positive when moving westward)

$$-m_e u_e^2/r = -eE+eu_eH.$$

Adding the two sets

$$m_iu_i^2+m_eu_e^2 \fallingdotseq m_iu_i^2 = reH(u_i-u_e). \quad .. \quad .. \quad (2)$$

Thus $u_i > u_e$ (though, as Chapman and Ferraro have shown, $u_i-u_e \ll u_i, u_e$). Hence, a westward electric current is necessary *on both sides of the earth* to allow the solar stream to flow in a curved path. A 'current-carrying-ring' is thus formed round the earth-dipole in the equatorial plane. For stability the 'ring' will be radially polarized.

If A be the (mean) cross-section of this 'ring', then the magnitude of the current is given by

$$i = Ne(u_i-u_e)A. .. \quad .. \quad .. \quad .. \quad (3)$$

This ring-current will produce a (negative) field at the earth of magnitude ΔH given by

$$\Delta H = 2\pi i/r. .. \quad .. \quad .. \quad .. \quad .. \quad (4)$$

Combining Eqs. (1), (2), (3) and (4) we get

$$A = 2z^5a^2\Delta H/0{\cdot}33.$$

If we take $z = 5{\cdot}5$ and $\Delta H = 50\gamma$ as the observed average decrease in H in the main phase of a magnetic storm, we have

$$A = 15a^2$$

If, for simplicity, we assume the cross-section to be circular of radius b, then $\pi b^2 = 15a^2$, or, $b \fallingdotseq 2a$, i.e. the radius of the cross-section of the ring is about twice that of the earth. This ring-current, which is a necessary

accompaniment of the steady state of the flow of the solar stream past the earth-dipole, remains, by virtue of its electrokinetic momentum, long after the stream has ceased to flow. (The stream flows only for about one or two days.)

(v) *Width of the auroral zone.*—It has already been said that the auroral zone on the earth is connected with the ring-current by the magnetic lines of force. It is found that the width ($\delta\theta_1$) of the annular terrestrial zone cut by all the lines of force which pass through the current-ring (of width δr in the equatorial plane) is equal to the observed width of the auroral zone. From the geometry of the dipole field we have

$$\delta r = -\frac{2a \cos \theta_1}{\sin^3 \theta_1} \delta\theta_1.$$

Taking $\delta r = 2b = 4a = 2{\cdot}6 \times 10^9$ cm., we get $\delta\theta_1 = 6°$. This is approximately the observed width of the auroral zone.

(vi) *Production of auroral phenomena.*—The immediate source of the high energy charged particles which cause auroral phenomena is the ring-current.

As already mentioned, for the stability of the ring, there will be radial polarization field between its inner and outer edges. The polarization on the surface of the ring is, however, (as shown by Chapman and Ferraro) unstable—charges of either sign being ejected from the surface along the lines of force. As the magnitude of the polarization field is of the order uH, (10^5 e.m.u.), the expellent force is comparable with (but less than) euH. The ejected particles have thus been impelled by an electric field of the order 10^{-3} volt /cm. extending over a distance of the order two earth-radii or 10^9 cm. Hence, when they reach the earth's atmosphere, after travelling along the guiding paths provided by the terrestrial magnetic lines of force, they will have sufficient penetrating power to produce the auroral phenomena (luminescence) and also to enhance the ionization and conductivity in the ionosphere, specially in the lowest regions.

(vii) *The S_D current system.*—From the preceding paragraph we see that the flow of the charged particles will produce closed systems of current sheets as follows: From the ring-current to the auroral zones (along the conducting paths provided by the lines of force), across the auroral zone (at about 90 km. level) and then back to the ring. This current is large, but since it flows in closed sheets above the ground it has small direct terrestrial magnetic influence.

However, since the inner and outer edges of the auroral zone are bombarded by particles of opposite signs (as the inner and outer edges of the ring-current, from which these particles are derived, are oppositely charged) there will be a potential difference between the two edges of the zone. According to calculations, the magnitude of this potential difference is somewhat greater than 10^5 volts, so that there is a meridional field of about 10^{-3} volt/cm. across the auroral belt. But, in the auroral belt there is also a nearly vertical magnetic field. Hence, under the influence

of this crossed electric and magnetic field, the ions in the auroral belt will drift at right angles to both, the drift velocity being

$$\frac{Ea}{H}\frac{\omega^2}{\omega^2+\nu^2}, \text{ where } \omega\left(=\frac{He}{m}\right)$$

is the ionic gyro-frequency and ν the ionic collision frequency.

Now, in the main phase of the storm, matter is streaming past the forbidden space, westward on the morning side and eastward on the evening side. The resulting polarization field in the 'ring' will be 'transferred' to the auroral zone in such a way that the ions there will drift from the daylight to its dark side under the combined action of this transferred field and the local nearly vertical magnetic field.

If the positive and negative ions had the same mass and collision frequency this would result merely in a transfer of ionized air from the day to the night side of the zone. However, it is highly improbable that both kinds of ions have the same mobility. If the positives drift some 10–30% faster than the negatives then the observed magnetic and ionospheric disturbance variations can be readily explained. In these circumstances a Hall current of some half-million amperes will flow round the auroral zone, westward on the dawn and eastward on the sunset sides. This current must set up a positive polarization over the dark and a negative polarization over the sunlit side. This polarization field is communicated from the highly conducting auroral zone to the poorer conducting regions over the rest of the earth. The form and magnitude of the resulting world-wide (and polar cap) current systems is that required to explain the S_D magnetic variations. The associated vertical electron drifts seem adequate to account for the disturbance variations in the F_2-region of the ionosphere. In the later stages of the storm, when the solar stream has ceased to flow the mass momentum of material in the ring is directed westwards all round the ring. The polarization field 'transferred' from the ring to the auroral zone now causes ions to drift westwards all round the zone. This now produces (the positive ions moving faster) a local westward (Hall) current in all parts of the auroral zone, but no polarization field can be developed by this current: thus the S_D variation dies away, leaving only the world-wide D_{st} field from the ring current, and a localized D_{st} field in the auroral zone.

(viii) *Other effects—Concluding remarks.*—It has been indicated by Martyn how the theory may explain the 'bay' magnetic disturbances and also (since, according to the theory, the S_D current system flows in the ionosphere) the effects as observed in the F_2-region during magnetic disturbances, which cannot be accounted for in terms of direct bombardment by solar particles.

The Chapman-Ferraro-Martyn theory of auroral and magnetic disturbance phenomena as sketched above is still in the formative stage. The theory, however, is soundly based, and, when fully worked out, may be

expected to present a complete picture of the mechanism of the auroral and magnetic disturbance phenomena and of the associated ionospheric effects.

5. NEUTRAL SOLAR CORPUSCULAR STREAM THEORY (Contd.) (ALFVÉN)

(a) Introduction

In Sec. 2(*b*) we have seen how, according to Alfvén, neutral ionized streams, laterally polarized, may be emitted from the sun. The associated polarization electric field is of fundamental importance in Alfvén's theory. We consider the special case in which the magnetic equatorial planes of the sun and the earth coincide and the particles are moving in this plane.

The width of the polarized solar stream is many times the diameter of the earth. (It is to be remembered that the linear dimension of a disturbed region in the sun, e.g. a sunspot, is many times that of the earth). We are thus required to investigate the motion of charged particles of the stream under the influence of the constant electric field of the stream (the linear extent of which is many times that of the dipole-field of the earth) and the magnetic dipole field of the earth (which is totally enclosed in the above).

(b) Motion of the stream in the magnetic field of the earth

As already indicated [Sec. 2(*b*)] the speed of the particles and the polarization characteristics of this wide stream at the distance of the earth's orbit from the sun are controlled by the value of the sun's magnetic field at this distance. It is found that particles of initial energy 10^8 eV, will have, at this distance, (according to the relation $V = kH$) energy of 1000 eV; the polarization electric field will be of the order 10^{-5} volt/cm.; the number-density of the particles (estimated from certain considerations) is found to be quite small, 10^{-4} particle/cm.^{3}

As the particles enter the terrestrial field, the forward motion, because of the increase in H, is retarded according to the relation $u = cF/H$. The inhomogeneous magnetic field of the earth causes *inhomogeneity drift*, driving the negatives and positives spiralling eastward and westward respectively. Since the motions are in the directions in which the electric force is acting there is an increase in the energy of the particles. This process of energy increase of the particles on entry into the earth's field is reverse of the process of energy decrease on emergence from the solar field [see Sec. 2(*b*)].

The path of the drift motion of a particle in the equatorial plane is now easily calculated. In Fig. 18 are shown the drift paths of an originally parallel stream of electrons coming from the sun. It will be seen that there is a forbidden region round the earth-dipole (at the centre of the figure) within which the electrons do not enter. Further, the paths on the morning side (right) and on the evening side (left) are not symmetrically disposed with respect to each other. The paths on the evening side are

more crowded and approach closer to the dipole than those on the morning side. (Note: The electrons advance along the drift paths spiralling round

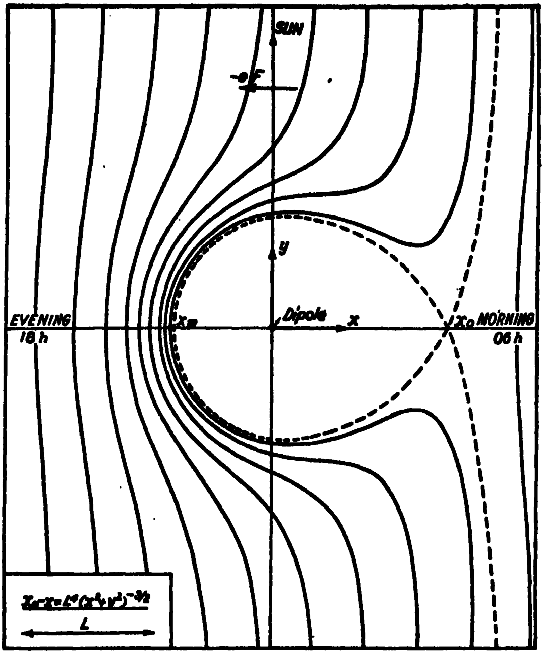


FIG. 18. Illustrating according to the theory of Alfvén the drift paths of an originally parallel stream of electrons coming from the sun and invading the field of the earth-dipole. [*Erratum.* In the figure, for x_0 (on the x-axis), read x_a.]

the magnetic lines of force, which run perpendicular to the plane of the paper. This is not shown in the figure.)

To calculate the path of an electron let the X-axis be in the direction of the electric field F and the Y-axis point towards the sun (see Fig. 18) and the origin be at the earth-dipole. The energy of the electron at a point (x, y) is given by

$$V = F(x_0 - x),$$

where x_0 is a constant giving the position of the electron relative to the Y-axis when it is at a very great distance from the dipole. The intensity of the magnetic field H at (x, y) is

$$H = a(x^2+y^2)^{-\frac{3}{2}},$$

where a is the dipole moment of the earth. Combining these two relations with the condition $V = kH$ we obtain the expression for the drift path of the electron characterized by the constant x_0 in the earth's magnetic field:

$$x_0 - x = L^4/r^3$$

where $$L = (ka/F)^{\frac{1}{4}}$$

and $$r = (x^2+y^2)^{\frac{1}{2}}.$$

It is seen from the figure that the stream of particles is divided into two branches, one passing by the morning side ($x_0 > x_D$) and the other by the evening side ($x_0 < x_D$) where x_D is equal to 1·76 L. The closest approach of the morning branch x_s is 1·32 L and that of the evening branch x_m is $-0{\cdot}74$ L. Between the two branches there is a forbidden region bounded by the dashed line in the figure. The size of this region is evidently determined by the value of L, which, in its turn, depends of k, F and a.

The drift paths of the positive particles, if they are of the same mass as the electrons, are just the mirror images of those of the electrons, there being crowding and closer approach to the dipole on the morning instead of on the evening side. For heavier particles, the curves are, however, 'diminished' in size. It can be shown that the forbidden region for such particles (for all reasonable values of mass and energy) lies entirely within that for electrons. It, therefore, follows that the electrons and positive ions in the portion of the neutral polarized solar stream as invade the dipole field of the earth, so divide themselves in their forward motion that a positive space charge (inside the electron-forbidden hollow) is formed on the morning side and a negative space charge (because the electrons are deprived of their positive companions) on the evening side.

(c) Magnetic disturbance and auroral phenomena

We can now follow the sequence of events described above to see how they may produce the main features of the magnetic disturbance and the auroral phenomena.

As the stream approaches the terrestrial magnetic field and as the positives and the negatives become differentiated and the stream loses its neutral character, magnetic effects begin to be felt on the earth. This is the beginning or the sudden commencement stage of a magnetic storm. As the stream invades deeper into the magnetic field with the electrons sweeping round the earth in the equatorial plane (in the eastward direction) we have the initial phase of the storm. When the forbidden region is fully formed, the motions of the particles become more stabilized and begin to give a maximum disturbance of the magnetic field at the earth. We now have the main phase of the storm. From Fig. 18 it is seen that the paths of the electrons (symmetrical about the 06^h—18^h line) form a sort of ring current—rather eccentric—in the equatorial plane in the westward direction. This part of the current system is responsible for the main part of the world-wide equatorial disturbance.

The positives differentiated from the negatives invade the negative-forbidden region and produce positive space charge near the boundary on the day side. The negatives going to the night side, being deprived of the neutralizing positives, produce negative space charge on the night side. These two space charges are connected with the earth round the poles by the magnetic lines of force which provide highly conducting paths (see Sec. 3). Hence, the two space charges neutralize themselves by being transported to the top of the atmospheres round the poles and discharging

themselves partly along the auroral zone and partly along the polar cap. The current system so produced causes the polar magnetic disturbances. The auroral zone may obviously be identified with the 'projection-line', i.e., the line which is a projection on the upper atmosphere of the boundary of the negative-forbidden region along the magnetic lines of force.

The auroral phenomena are produced by bombardment of air molecules by electrons coming down from above along the lines of force.

Fig. 19, constructed by Cowling [29] from Alfvén's description, depicts the main features of the theory for the region round the earth.

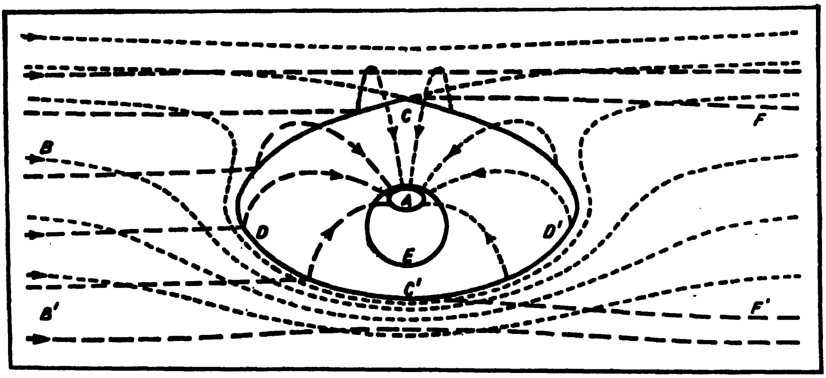


FIG. 19. A perspective drawing showing according to the theory of Alfvén, the paths of ions and electrons of the neutral solar stream as they invade the earth-magnet. Only the paths of the charges reaching the north auroral zone are shown for simplicity. The thick and the thin dashes indicate the paths of the ions and electrons respectively. E—earth; curve A—north auroral zone; BB'—ions and electrons approaching earth in equatorial plane; $CDC'D'$—boundary of forbidden region; CF and $C'F'$ are boundaries of 'shadow' of earth where no ions are present (Illustrated after Cowling.)

Alfven has also subjected his theory to quantitative tests and has shown that most of the detailed features of auroral and terrestrial magnetic disturbance phenomena may be deduced as consequences of this theory. For example, the normal polar distance of the auroral zone is of the same order as that of the average of the projection-curve. Further, from the theoretical expression for the size of the auroral curve, it is found that the distance varies at the eighth root of ka/F. Hence k and F can vary (as they naturally will) over wide ranges without affecting the polar distance greatly.

Amongst other features that are explained are, occurrence of the more brilliant auroras before midnight, the diurnal variation of the auroral frequency curve (two maxima; see Fig. 8, Chapter VIII), the fact that the currents producing the polar disturbances flow eastward in the afternoon side and westward in the morning side and the directions of the auroral arcs at different times of the day.

Further, experiments similar to those of Birkeland and Brüche have been performed by Malmfors to test the theory [27]. A magnetized steel sphere ('Terrella') representing the earth was placed in a vacuum chamber (0·01–0·001 mm. Hg) and an electric field to represent the polarization field in the ionized stream was applied. The 'Terrella' was covered with a fluorescent powder. On introducing the ionized particles of the stream eccentric luminous rings were produced round the poles. According to Malmfors if the rings are scaled up to dimensions as found in Nature, then they would correspond, even quantitatively, with the auroral zones.

(*d*) Criticisms of the theory

The whole of Alfvén's theory is, of course, based on the effect of the supposed general magnetic field of the sun. Hence if this field is non-existent, (as is now believed [17]), there is no basis for the theory.

Apart from the above, the proposed behaviour of the particles of the neutral ionized stream, on entering the earth's magnetic field, has been subjected to criticisms by Cowling [29] and by Chapman [30]. According to Cowling the electrostatic field of the space and surface charge distributions as would be produced by the supposed motions of the charged particles, would radically alter these motions and, thus, destroy the apparent agreements with observations. Cowling concludes 'that diamagnetic repulsion prevents the particles from reaching the earth's atmosphere, and that electrostatic forces prevent more than a slight separation of ions from electrons. It seems that a stream of ionized particles can reach the earth only if either the particles possess such large energies that we cannot divide their motion roughly into a spiral motion along a line of force and a relatively slow drift across the lines of force, or if the stream is so dense that the leading particles are able to shield the body of the stream from diamagnetic repulsion' (*loc. cit.*, p. 212).

Ferraro and Chapman also come to a similar conclusion from a consideration of the 'terrestrial' part of the theory. Alfvén's inpulsion process of the solar stream (assuming that there is a general magnetic field of the sun) has also been examined by Chapman. According to Chapman [30] 'before the stream reached a distance of 10 solar-radii from the sun, the lateral drift would have completely separated the positive and negative charges, which would be quickly dispersed. Thus a stream of the supposed kind could not reach the earth.'

6. ULTRAVIOLET LIGHT THEORY (PAULSEN-HULBURT-MARIS)

The interesting suggestion that auroral phenomena and geomagnetic disturbances may be caused by charged corpuscles, not of solar but of terrestrial origin produced by the action of ultraviolet rays of the sun on the upper atmosphere, seems to have been first made by Paulsen [31, 32]. A bold attempt has been made by Hulburt and Maris to build a

complete theory of auroras and magnetic storms round this idea. The theory has, however, been modified more than once in view of criticisms made by Chapman and is still open to grave objections. Nevertheless, the theory contains many stimulating and interesting ideas and a brief account of it is given below.

The theory as originally proposed by Hulburt [33] assumed that the molecules in the rarefied atmosphere above 200 km. are ionized by the solar rays and the ion pairs thus formed are guided by the lines of force of the earth's magnetic field towards the poles. The ions—positive and negative—formed in the northern hemisphere go towards the north pole and those formed in the southern hemisphere go towards the south pole. It was pointed out, however, by Chapman [34] that the magnetic lines of force, even at heights of 1,000 km. above the equator, meet the surface of the earth at magnetic latitude of about 20°. Hence the charged particles which are guided by the lines of force cannot reach the auroral belts. In fact, the lines of force which meet the earth in auroral latitudes (about 65°) pass tens of thousands of kilometres above the equator. The density of the atmosphere at such high levels being quite inappreciable, the theory in its original form had to be discarded. The subject was, however, further pursued in a series of papers by Maris and Hulburt [35, 36, 37] and a modified theory was put forward.

The essentials of this new theory are as follows. The exosphere or the fringe region of the atmosphere [Chapter I, Sec. 5(*c*)] where the atmospheric particles experience negligible collisions and dance up and down by receiving impacts from the dense atmosphere below, is assumed to extend from above 300–400 km. The number of molecules in the exosphere is estimated to be 10^{16} molecules per square cm. column and they experience 10^{14} impacts per second with the molecules in the denser region (300–400 km.) below. Assuming a temperature of 1,000°K in this latter region the velocities imparted to the particles are found to be of the order of 1·7 km./sec. These velocities take the particles to heights of 2,000 to 3,000 km. In order that the particles may rise to heights of 40,000 to 50,000 km.—the heights to which the magnetic lines of force meeting the earth at the auroral belt rise above the equator—it is assumed that the atmosphere at 300–400 km. level contains excited molecules and that one in a million of the 10^{14} impacts per second which the molecules in the fringe region experience (with the molecules in the 300–400 km. level) is of a superelastic kind. Justification is sought for this assumption on the ground that many lines or bands of aurora and of night air-glow originate from metastable levels of O and of N_2. Thus the 10^{14} collisions per second produce 10^8 high speed particles per second. Of these, one half fly downward and are lost in the dense atmosphere below and the other half fly upwards. If the velocity is assumed to be 10 km./sec. or more the particles will attain heights of 30,000 to 50,000 km. in 3 hours.

The next step in the theory is to consider the ionization of these high flying particles. It is shown that the probability of ionization is such that

the particles are ionized after they have reached the topmost parts of their trajectories. This is necessary as otherwise if they are ionized while on their way up, the ions formed would be deflected down by the earth's field. The ion pairs formed at the top (30,000–50,000 km.) are then guided towards the polar regions spiralling round the lines of force and take 6 to 14 hours to fall into the 25° zone round the magnetic poles where they give rise to the auroral phenomena. Since the ions are supposed to be carried round by the lines of force, those formed during noon drop to the polar regions at 9 p.m. or later and produce the auroral phenomena.

Great magnetic storms and associated auroral displays are attributed to sudden outbursts of ultraviolet radiation from the sun. The assumption undoubtedly receives support from the phenomenon of radio fade-out [Chapter VI, Sec. 13(*d*)]. In regard to the mechanism of such emissions Maris and Hulburt consider the following:—'Imagine that by some unknown process a part, say 1/10,000th of the solar surface, is removed and the interior at a temperature of, say, 30,000°K is exposed to emit black body radiation. The total solar radiation is then increased by 63 per cent. This increase, however, is not uniformly distributed over the entire wavelength range but is enormously concentrated in the extreme ultraviolet. While the solar constant is increased by 1 per cent only, the energy in the wavelength ranges λ 500 to λ 1,000 is increased 10^5 times.

The intense blast of ultraviolet radiation causes a sudden increase in the ionization and, according to the authors [38] of the theory 'the first phase of the average world-wide magnetic storm is attributed to the sudden increase in the eastward ion drift current which girdles the earth caused by an increase in the long free path ions produced by a solar ultraviolet flare. The second phase of the storm comes about from the heating of the high atmosphere by the flare. The atmosphere expands and the outward movement of the ionized regions across the earth's magnetic field gives rise to a westward current in the high atmosphere flowing around the earth. The movement also decreases the long free path ions and increases the short free path ions; this prolongs the westward current.'

As already mentioned the theory is open to many objections. Some of these are discussed below.

It has been shown [39] that no super-elastic collision process can be envisaged in the 300–400 km. region by which particles with speed high enough to shoot up to 30,000–50,000 km. can be produced. The greatest heights attained by such particles are 9,500 km. to 14,000 km. And, for reasons already mentioned, ions formed at such heights cannot be guided by the lines of force to fall into the auroral belt.

According to Ta-You Wu [40] the time taken for ionization of the upgoing particles by the solar ultraviolet rays is not 3 hours as assumed by Maris and Hulburt, but thousand times or more than that.

A fundamental difficulty has been pointed out by Chapman. He draws attention to the fact that since terrestrial magnetic field does not increase the energy of the particles, the velocity of the ions is the same as

that assumed at the starting point, namely about 10·5 km./sec. This speed is totally inadequate for penetration of the particles into the dense atmosphere down to the auroral levels.

For explaining the 27-day recurrence tendency of magnetic storms it is necessary to assume that the solar outbursts send out the ultraviolet radiation in the form of narrow beam. It is, however, difficult to imagine how a bright patch on the sun's surface would send out a narrow beam instead of a wide angle flare. One has to suppose that the radiation comes out of deep holes or is in some way focussed so as to be in the form of a beam. In order to avoid this difficulty Hulburt postulates that the ultraviolet emissions may come out as wide angle puffs. And, to account for the recurrence tendency the *ad hoc* assumption is made that the flare puffs out periodically with a recurrence period of 27 days due to pulsations of some sort.

Observations on the simultaneity of radio fade-out, incidence of geomagnetic disturbance and appearance of bright solar patches [Chapter VI, Sec. 13(*d*)] also go against the ultraviolet light theory. It has been shown by McNish [41] that the magnetic disturbance is not of the storm type, but is merely an enhanced dynamo effect—i.e., an enhancement of the quiet day current system (S_q) due to increased conductivity of the ionospheric regions produced by intensification of ionization.

Besides the above, the explanations of the other features of terrestrial magnetic variations and of auroral phenomena offered by Maris and Hulburt on the basis of their ultraviolet light theory have been subjected to serious criticisms by Chapman. For details the reader is referred to the original papers. (An interesting account of the theory and the objections to it are to be found in Chapter XXV of *Geomagnetism* by Chapman and Bartels.)

CHAPTER X

LIGHTS FROM THE NIGHT SKY

1. INTRODUCTION

To a casual observer the only light received in a dark moonless night, in an open field away from city lights, appears to be that coming from the stars. Measurements show that under such conditions the intensity of illumination near the ground is, on the average, equal to that produced by a small surface held normally to the rays of a standard candle placed at a distance of 57 metres [1]. (The word 'average' is used advisedly because, as will be presently seen, the intensity is strongly variable. See for instance references [2, 3].) Assuming that this light is received from the stars alone, this would be equivalent to a uniform distribution of 0·045 star of photographic magnitude unity for every square degree. Careful computations show, however, that the total light sent out by all the stars both telescopic and observable in the polar region is equivalent to that received from a uniform distribution of only 0·009 star of photographic magnitude unity for every square degree of the sky [4].

It is thus obvious that the portions of the sky which are entirely devoid of stars must be continually sending out light and contributing to the major portion of the illumination observed in a moonless night. Investigations have shown that this light from the night sky is composed of lights from various sources. In the first place, there is the luminosity of the sky due to scattering of star light by the air molecules in the same way as the scattering of sunlight during daytime to which the blue of the sky is due. Secondly, the sky is illuminated by the zodiacal light about which we shall speak in greater detail later. Thirdly, there is the so-called galactic light which is due to the scattering of light of nebulae and of stars by the diffuse matter in the interstellar space. Fourthly, there is the light due to the self-luminescence of the upper atmospheric gases or the '*air-glow*'. This is the most important of all the sources of the light and will form the main subject-matter of this Chapter. The light is also sometimes called *permanent aurora* because of some common features between the spectrum of this light and that of the polar aurora [see Chapter VIII, Sec. 6*a*]. The adjective 'permanent' is used since, unlike the polar aurora which shines only occasionally, the 'air-glow' is permanently present. (*Note:* The useful word 'air-glow' is due to Elvey [5].)

Finally, to complete the list, there is luminosity due to scattering by the atmosphere of the light from all the above sources (the scattering of star light has already been mentioned).

The various sources and the average percentage of their contributions in the visible region are enumerated in Table I.

TABLE I

Star light, direct and scattered	30 per cent.
Zodiacal light	15 ,,
Galactic light	5 ,,
Luminescence of the night sky (Air-glow)	40 ,,
Scattered light from the last three sources	10 ,,

Of the relative intensities of the different sources given in Table I, those due to the zodiacal light and to the night air-glow are subject to considerable fluctuations. The figures in Table I give only the average distribution.

Before discussing the night air-glow, which, as already mentioned, is the most important of the night sky radiations, we will describe the zodiacal light and the galactic light.

2. ZODIACAL LIGHT AND GEGENSCHEIN

(a) Zodiacal light

(i) *General characteristics.*—On clear moonless evenings a tongue or cone of faint light is often seen to rise above the horizon in the west after the disappearance of the twilight (Fig. 1). The phenomenon appears

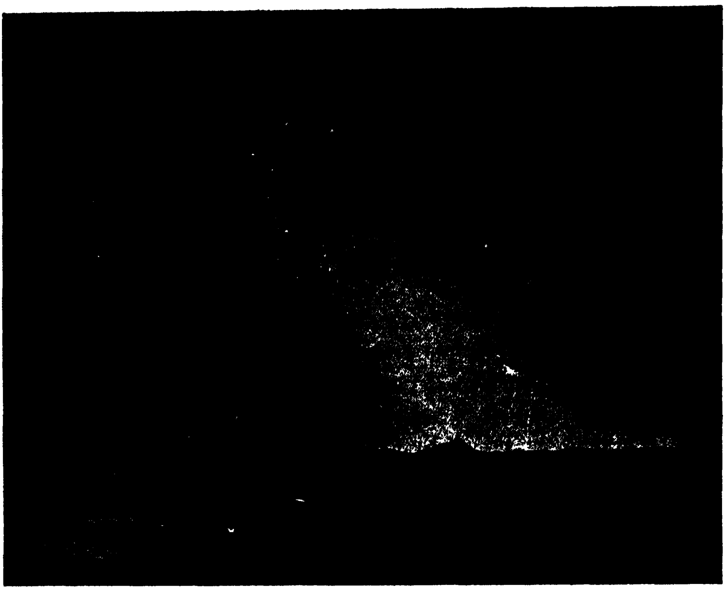
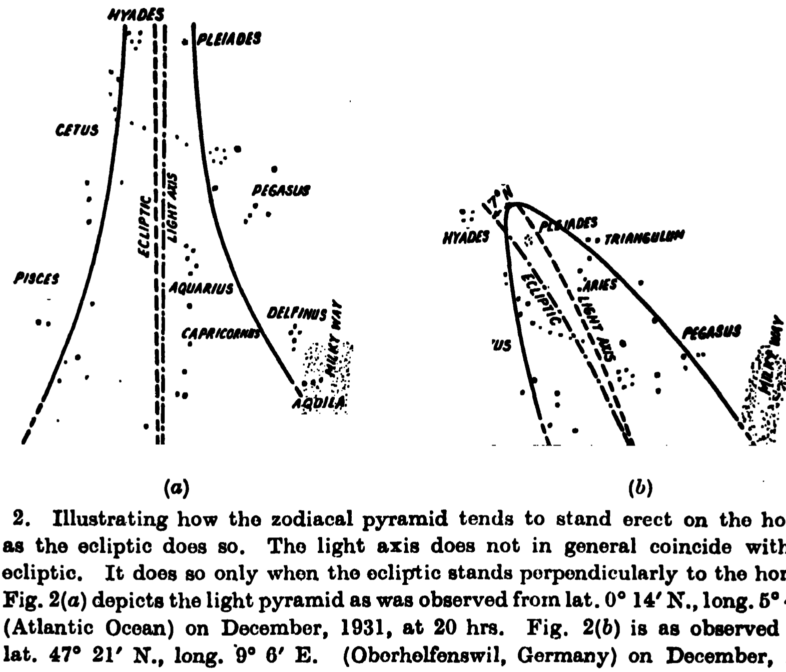


FIG. 1. Zodiacal light. (*Photograph by Rudaux, taken on April 26, 1913.*)

distinctly in middle and lower latitudes. The colour of the light cone is reddish yellow to white; its intensity is comparable to that of the milky way, being brighter near the base and growing fainter at higher altitudes. The main portion of the light pyramid lies in the zodiacal belt; hence the

name zodiacal light. The base of the cone extends on the horizon 15° to 20° on either side of the point of intersection of the ecliptic with the horizon. Upwards, it rises 50° to 60° from the base. A similar cone of light is also seen in the east before the incidence of the first morning rays. The former is called the west zodiacal light and the latter, the east zodiacal light.

The pyramid stands erect on the horizon when the zodiacal belt cuts the horizon perpendicularly. This occurs during the equinoxes in the tropics—in the evening for the spring equinox and in the morning for the autumnal equinox. In the equatorial regions the ecliptic always stands more or less erect, being more specially so at sunrise and sunset hours during the solstices. Observed from higher latitudes the zodiacal pyramid is inclined to the horizon as is the zodiacal belt, and the inclination varies according to the season and the hour of observation [Figs. 2(*a*) and 2(*b*)].



(*a*) (*b*)

FIG. 2. Illustrating how the zodiacal pyramid tends to stand erect on the horizon as the ecliptic does so. The light axis does not in general coincide with the ecliptic. It does so only when the ecliptic stands perpendicularly to the horizon. Fig. 2(*a*) depicts the light pyramid as was observed from lat. 0° 14′ N., long. 5° 4′ W. (Atlantic Ocean) on December, 1931, at 20 hrs. Fig. 2(*b*) is as observed from lat. 47° 21′ N., long. 9° 6′ E. (Oberhelfenswil, Germany) on December, 1922, 23 hrs. 30 mins. (Both the drawings are after Schmid.)

As the sun sinks below the horizon the pyramid sinks with it but the light is usually discernible on the horizon till the angular distance of the sun reaches 90°. In low latitudes and in very clear weather, the zodiacal light may be seen to extend up to 180° from the sun. The west and the east zodiacal lights then join together forming a continuous bridge of light close to the zodiacal belt. This is called the *zodiacal band*. Its width and brightness decrease with angular distance from the sun. At 90° it may be about 20° wide and at 150° about 10°.

According to some authors the zodiacal light really extends over the whole sky though, of course, it is strongest along the light bridge in the zodiacal belt. Such extension of zodiacal light can be observed only from fairly high altitudes when the atmosphere is quite free from pollution.

The following description of the changes of zodiacal light in course of a year in moderately high latitudes is taken from Schmidt who has made a careful study of the phenomenon [6]: 'The west zodiacal light begins to appear by the end of September and in November fully develops the character of zodiacal light. The cone of light rises out of the south-west horizon as the ecliptic rises and it reaches the highest point in the month of January. It then sinks gradually and is lost in the summer night glow in the month of May. The sky is free from zodiacal light in the month of June. By the end of July we see the first appearance of morning zodiacal light shortly after midnight. At the beginning the cone of light is strongly inclined to the horizon like the ecliptic. With the approach of autumn the inclination gradually changes and in the month of November the cone of light becomes almost vertical. It then begins to incline more and more towards the south with increasing inclination of the ecliptic and disappears in the south-eastern sky towards the middle of March.' (For general descriptions of zodiacal light, see references [6, 7, 8].)

Though the main portion of the light pyramid lies in the zodiacal belt, the axis of the pyramid does not in general coincide with the ecliptic. It does so only when the ecliptic stands erect on the horizon, that is, when the observer is situated in the plane of the ecliptic. For inclined position of the ecliptic, the light axis is deviated from it; the greater the inclination the greater is the deviation. Observed from northern latitudes the deviation is to the north and from southern latitudes to the south.

According to Schmidt, the change in the position of the light axis with reference to the ecliptic may also be observed in course of a single night because, in general, the motion of the observer due to diurnal motion of the earth on its axis, transports him towards or away from the plane of the ecliptic. Under favourable conditions, as for instance, when the west

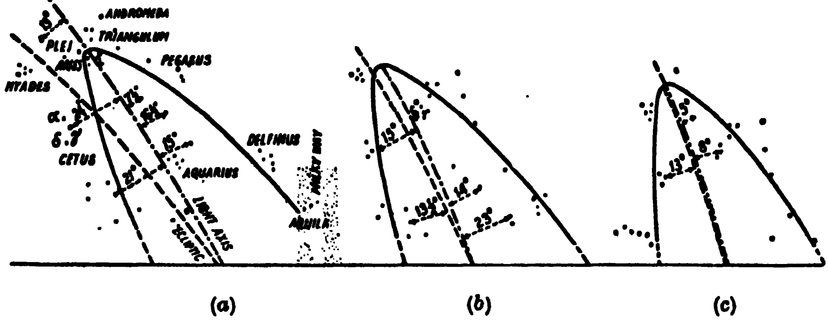


FIG. 3. Illustrating how the position of the light axis changes in course of a single night. Observations made at Oberhelfenswil (47° 21′ N., 9° 6′ E.) on 9 Jan., 1934. Fig. (a) for 19 hrs., (b) for 22 hrs., and (c) for 24 hrs. (After Schmidt.)

zodiacal light is observed from high northern latitudes in winter months, the change of position of the light pyramid (against the starry background) might be considerable in course of a few hours (Fig. 3). The whole light cone not only tends to stand erect with the progress of night (as the zodiacal belt does so) but also appears to move sideways, so that the light axis which was originally to the right of the ecliptic crosses over slightly to the left [9]. These movements of the zodiacal light are called its nightly motion. (It is to be mentioned, however, that Hoffmeister [22] who has made extensive observations from South Africa and also from other places, has questioned the reality of these nightly motions as mentioned by Schmidt.)

(ii) *Intensity variation.*—It has already been mentioned that the intensity of the zodiacal light is comparable to that of the milky way. Measurements at Montpellier, France, show that the intensity is about 2·4 times that of the night air-glow near the region of the pole star. The intensity varies considerably from night to night and also from hour to hour in a particular night. At times the light may be almost completely

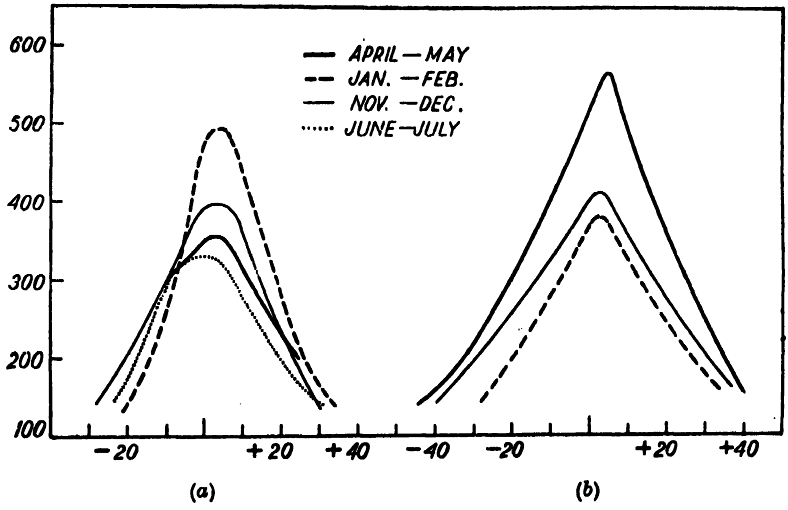


FIG. 4. Sections across (*a*) the morning, and (*b*) evening zodiacal light at longitudes 300° and 60° respectively from the sun, for different seasons of the year as observed at Texas. It will be noticed that the evening zodiacal light has the greatest intensity in April-May and least intensity in Jan.-Feb. The reverse is the case for the morning zodiacal light. The ordinate gives the intensity in units of tenth photographic magnitude per square degree of the sky; the abscissa marks the latitudes. (After Elvey and Roach.)

extinct. It appears that some at least of such large variations are coincident with the incidence of magnetic storms [10]. Some authors believe that the rapid variations in the intensity of zodiacal light might also be coincident with solar outbursts associated with bright eruptions on the sun's surface [11]. Meteoric activities also have been suggested by some to be associated with the variation [12, 32]. The meteors responsible, however, are the

larger ones, as the absence of Rayleigh scattering in the spectrum suggests [15].

Fig. 4 depicts the seasonal as well as the latitudinal variations of the evening and morning zodiacal light after quantitative measurements of Elvey and Roach, made at Texas with a photoelectric recording photometer [13]. The ordinates give the intensities in units of the tenth photographic magnitude per square degree of the sky. It will be seen that in November-December, the evening zodiacal light is rather inconspicuous; for the morning zodiacal light, the gradient of intensity is steeper. Further, for the evening zodiacal light, the maximum intensity is found to occur in April-May, while the minimum occurs in January-February. The reverse is the case for the morning zodiacal light. It may be mentioned, however, that Japanese observers have found a maximum in January for the evening light and in October for the morning light [12].

The phenomenon of *zodiacal twilight* (as distinguished from ordinary twilight) may be mentioned here. This is observed till the sun goes 25° to 30° below the horizon and is characterized by the strengthening of the emission lines of the night air-glow, but not of its continuous spectrum (*vide infra*). The strengthening depends essentially on the inclination of the ecliptic to the horizon. When the ecliptic tends to be vertical to the horizon, the zodiacal twilight becomes intense and produces considerable widening of the light pyramid specially near the base. This widening can be seen, for example, in Fig. 2(*a*) and in Fig. 4.

(iii) *The spectrum.*—The spectrum of the zodiacal light (like the spectrum of the night air-glow; see Sec. 4*c*), consists of lines and bands superposed on a continuous background. So far as the former is concerned they are the same as in the night air-glow, only somewhat more intense [16, 17, 18]. (No lines or bands as are not observed in the night air-glow have been detected.) According to the careful photometric measurements of Karimov [19] (made in 1947-48 at the Gornaya Mountain Astrophysical Observatory, 15 km. south of Alma Ata SSR) the intensities of the red and green oxygen lines on the axis of the zodiacal cone were found to be 1·40 and 1·53 times respectively greater than those in the night air-glow. Measurements were also made on the yellow sodium lines. But the results were erratic. On some occasions there was three-fold increase of intensity. On other occasions the lines were completely absent from the night sky as also from the zodiacal cone.

In contrast to the above, the continuous part of the spectrum is quite different from that of the night air-glow. (For energy distribution in night air-glow continuous spectrum see Fig. 12.) On the other hand, a marked similarity with the solar spectrum is observed. According to measurements of Karimov referred to above, the colour-temperature of the zodiacal light is about 6000°K and the colour-index about 0·53, corresponding to the spectral Class *G*, that is, to that of the sun [17]. (These results were obtained after the contribution—about 25%—from the night air-glow continuous spectrum had been removed.) It may also be mentioned

that some observers have identified the Fraunhofer absorption lines in the zodiacal light spectrum [14]. These facts have important bearing on the theory of zodiacal light which we will presently discuss.

(iv) *Polarization.*—Comparative study of the polarization of light from regions of the sky in the zodiacal cone and from regions outside it has been made by Dufay and Déjardin [20, 1]. It is found that while 15 per cent of the light from the zodiacal cone is polarized, only 2 to 4 per cent of the light from the night sky is so. The planes of polarization of both are found to pass constantly through the sun. The results obtained by Dufay are shown in Fig. 5. The ordinate represents the depolarization factor which is the ratio of the minimum to the maximum of light transmitted through the analyzing nicol when it is rotated in the line of sight. (The region of the night sky observed was in the meridian towards the north, while that for the zodiacal cone was necessarily in the region of the ecliptic.) The upper curve is for night air-glow and the lower curve for the zodiacal light. The stronger polarization of the zodiacal light is evident from the figure. It will be noticed that the general trend of its variation is similar to that of the night air-glow. It has, therefore, been suggested that the polarized component of the night air-glow is only an extension of the zodiacal light. [See also Sec. 5(*g*).]

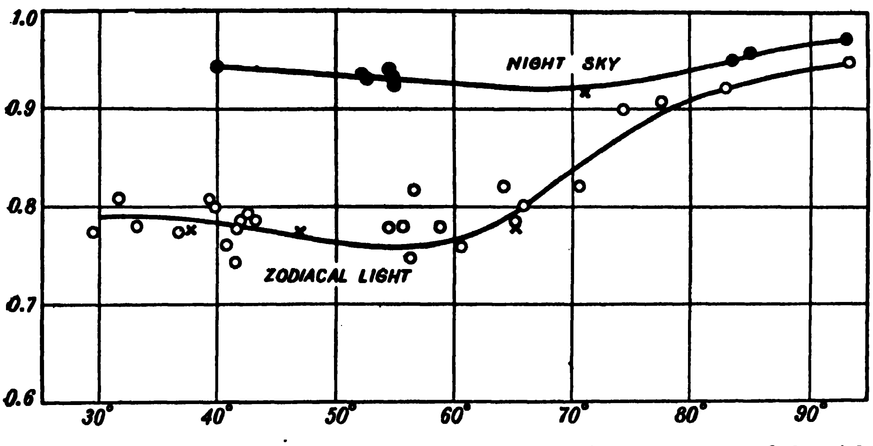


FIG. 5. Comparison of the polarization of the zodiacal light with that of the night air-glow. Note that the zodiacal light is much more strongly polarized than the night air-glow. Proportion of polarized light is equal to $(\rho-1)/(\rho+1)$ where ρ is the depolarization factor. See text. (After Dufay.)

(v) *Parallax.*—Attempts have been made to find out the distance at which the zodiacal light originates by measuring its parallax [21]. This has not been successful as no parallax could be detected. According to some authors this absence of parallax shows that the origin of zodiacal light and gegenschein must be at great distance from the earth beyond the orbit of the moon or even further away.

(*b*) Gegenschein and False Zodiacal Light

On the zodiacal band—which stretches across the sky along the ecliptic—a patch of delicate glow is observed in the region opposite to that occupied by the sun. On account of its being (nearly) opposite to the sun the glow is called *gegenschein* or *counter-glow*. The glow is rather faint in the early parts of the night and becomes distinct only towards midnight. It is elliptic in shape with its major axis lying nearly along the ecliptic. (There is some observable deviation from the ecliptic when the gegenschein is in the western part of the sky.) On account of its faintness it is difficult to assign any accurate dimension to the gegenschein. Generally, it covers a region of a few degrees—about 10° along the major axis and 6° along the minor axis, though, on favourable occasions, it may attain much larger size 40° by 10° [21*a*]. The centre of the patch is not exactly opposite to that of the sun but is slightly deviated westward from the anthelion.

The brightness of the gegenschein is on the average about 13% higher than the general brightness of the surrounding night air-glow. The brightness does not remain constant throughout the night. This may, however, be explained (apart from probable observational error) as due to fluctuations of the coefficient of transparency of the atmosphere. But, the brightness has also a strong seasonal variation; it is on the average about 40% brighter in the spring than in autumn. There is a marked correlation between these variations and the brightness of the night air-glow.

According to observations made by Karimov (as mentioned in Ref. 21*c*) the spectrum of gegenschein is closely similar to that of the night air-glow. The intensity of the line emissions in the gegenschein is increased by about 10–15%, but there is no perceptible strengthening of the continuous background.

It appears that the gegenschein oval has a definite nightly parallactic movement. This was first detected roughly by Astapovich [21*d*] by visual observations and later confirmed by Rozhkovski [21*b*] by more accurate photometric measurements of photographs made during the years 1947–49 at the Gornaya (Mountain) Astrophysical Observatory near the town Alma Ata, Kazzakh, SSR. Systems of isophotes of the counter-glow for a number of nights were prepared (after making due optical corrections and eliminating the contribution to the brightness from other sources, e.g. night sky air-glow, galactic light) which revealed a diurnal (24-hour) horizontal parallax amounting to about 3·5°

A phenomenon closely associated with the gegenschein is the so-called false zodiacal light (Fig. 6). It was first noticed by Divari on the glacier Tooyouk-Soo at an elevation of 3·5 km. and later by Fessenkov [21*c*] in the desert of Southern Balkhash region (Sary Ishik Otraou). In the early hours of the morning, some 2 to 2½ hours before sunrise, the gegenschein oval (when it is about 40°–45° above the western horizon) is distorted and seems to open out into a cone with its axis in the plane of the ecliptic and its tail pointing west [21*a*, 21*b*]. It thus has the appearance of the zodiacal

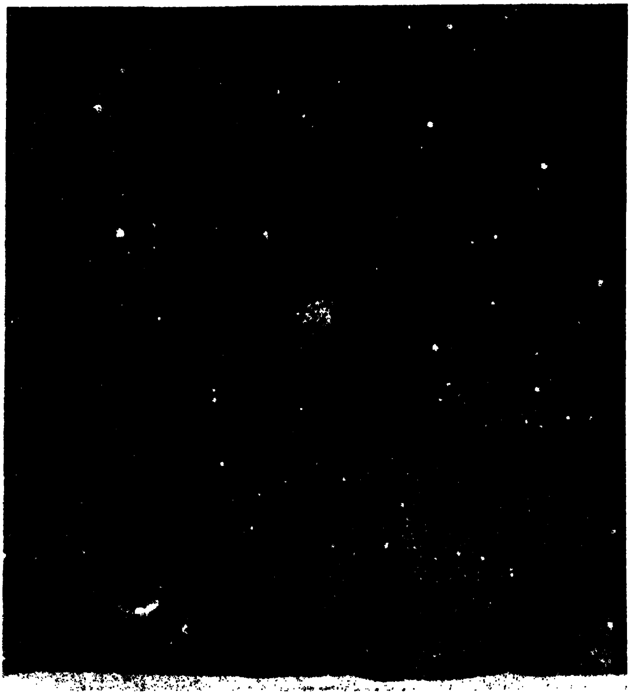


FIG. 6. False zodiacal light and gegenschein observed in the desert Sary Ishik Otraou, October 1948. *Note:* As a result of increased contrast in the reproduction, the gegenschein appears much brighter than in the original. (After Fessenkov.)

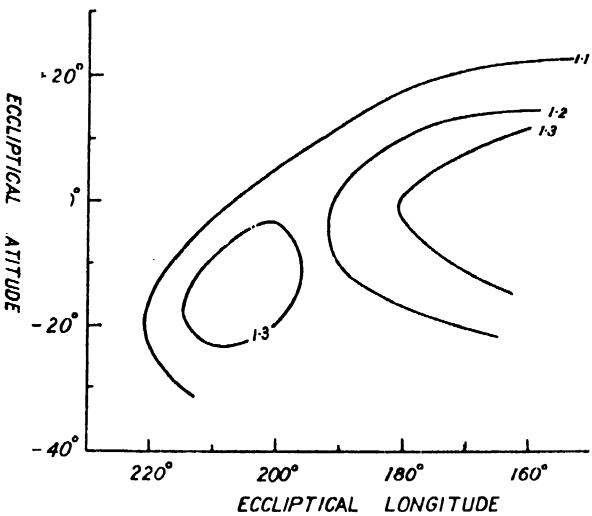


FIG. 7. Isophotes of the gegenschein at the stage when it is opening out to form false zodiacal light. Observed at Gornaya (Mountain) Observatory near Alma Ata, SSR. on October 25, 1946, at 3.33 a.m. (After Divari.)

pyramid and, hence, is called *false zodiacal light* to distinguish it from the former, which, in the early morning, is beginning to appear on the eastern horizon. The false zodiacal light also appears to be formed by conspicuous widening of the zodiacal band (which stretches across the sky along the ecliptic) in the later part of the night but is not to be confused with the same. Fig. 7 shows the isophotes of the gegenschein (distorted) as constructed by Divari from observations made at 3h. 33 min. on October 25, 1946, at the Gornaya Observatory. They clearly show how the oval has opened out to form the cone of the false zodiacal light.

(c) Theories

(i) *Introduction.*—There can be little doubt that the zodiacal light and its associated phenomena are due to light received from extensive cloud of material particles of some sort, illuminated by solar rays. Two hypotheses have been current regarding the location of the cloud in the solar system. According to one—the so-called *planetary theory*—the cloud belongs to the sun; according to the other—the *atmospheric theory*—it belongs to the earth. The earlier theorists attempted to explain both zodiacal light and gegenschein phenomena, exclusively with the help of the one or the other theory. According to contemporary workers, however, both planetary and atmospheric theories have to be invoked for a complete explanation of the whole group of phenomena associated with zodiacal light and gegenschein. The scattering cloud, according to both the theories, is of a very flat shape, the scattering particles being confined mostly in the plane of the ecliptic.

In what follows we will give brief accounts of both the planetary and the atmospheric theory indicating how far they are able to explain the zodiacal light and the gegenschein phenomena.

(ii) *Atmospheric theory (Schmidt).*—Of the several formulations of the atmospheric hypothesis, that by Schmidt deserves special mention [3]. Schmidt assumes that the earth is surrounded by a lens-shaped cloud (which may be an extension of the terrestrial atmosphere) with *its plane lying in the plane of the ecliptic.* The extension of the cloud is uncertain but it must be much beyond the distance at which the centrifugal force due to the earth's rotation equals its gravitational acceleration (about 6·6 earth-radii). The cloud lens participates in the motion of the earth round the sun but its plane always remains in the plane of the ecliptic.

Figs. 8(*a*) and (*b*) explain, after Schmidt, how the portion of the atmospheric dust lens illuminated by solar rays will appear to an observer as a pyramid of light [9]. In Fig. 8(*a*) the inner circle represents the earth's equator and the outer circle the periphery of the scattering lens. The illuminated portion of the lens is shown as hatched. The observer is situated at *B* and the trace of his horizon is shown by the tangent through *B*. The sections of the illuminated portion of the lens visible to the observer are shown by double hatching. It is obvious that the observer, looking

along the tangent, say along the line BA, will observe a light pyramid such as that shown in Fig. 8(b).

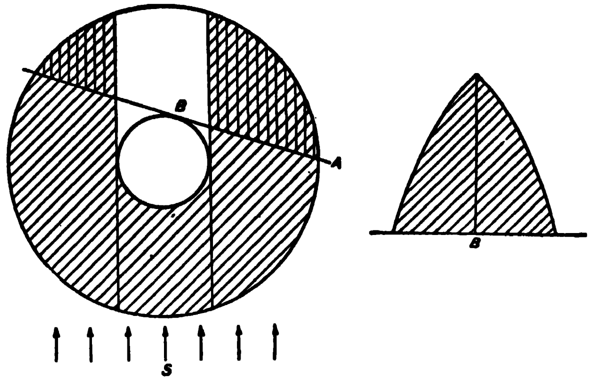


FIG. 8. Illustrating the formation of zodiacal light pyramid. The earth, inner circle of figure (a), is surrounded by a lens shaped cloud, the periphery of which is the outer circle. The portion of the lens illuminated by solar rays is shown shaded. The observer on earth is at B. Looking towards the horizon along the tangent through B he will see a portion of the illuminated lens (shown by double hatching). This will appear as an illuminated pyramid as in figure (b). (After Schmidt.)

Fig. 9 depicts the positions of the earth round the sun in the various seasons with the respective illuminated portions of the lens as

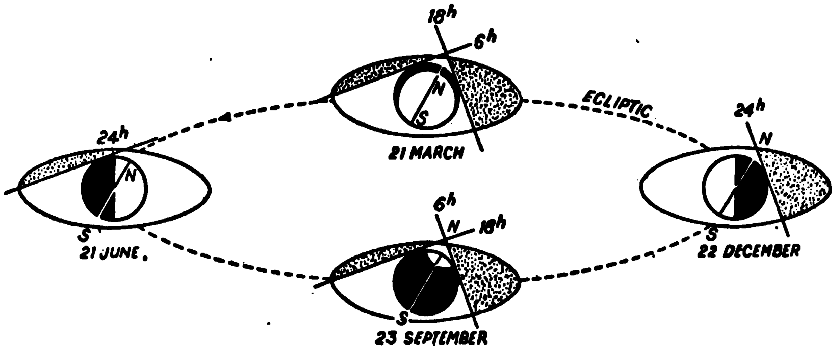


FIG. 9. Illustrating the seasonal variation of the intensity of zodiacal light as observed from middle latitude in the northern hemisphere. The earth is surrounded by a lens-shaped cloud of dust situated in the plane of the ecliptic. Since the intensity of the light depends upon the amount of matter illuminated by solar rays, it is evident that the morning zodiacal light will be prominent from the autumnal equinox to the winter solstice and the evening zodiacal light from the winter solstice to the vernal equinox. (After Schmidt.)

seen by an observer in middle latitudes in the northern hemisphere [9]. Since the intensity of the light pyramid depends upon the amount of matter illuminated by the solar rays, it is evident that for the positions of the

observer as stated above, the morning zodiacal light will be prominent from the autumnal equinox to the winter solstice and the evening zodiacal light from the winter solstice to the vernal equinox.

The seasonal variation of the tilt of the light pyramid is also understood if it is recalled that the axis of the pyramid always tends to coincide with the ecliptic.

Schmidt explains the nightly deviations of the light pyramid axis from the ecliptic as due to an extinction effect produced by the nightly motion of the observer due to the rotation of the earth. He further asserts that since the nightly deviations are in directions opposite to those expected from parallax, any parallax effect will be masked thereby. It is to be stressed in this connection that since the zodiacal pyramid appears to change its position (against the starry background) with the nightly change of position of the observer—be that due to extinction or to parallax—it is evident that the scattering cloud system cannot be very far from the observer and that it presumably lies within the solar system. (Hoffmeister's failure to observe the nightly movement should, however, be recalled in this connection. *Vide supra.*)

Schmidt also explains the midnight gegenschein as follows. The region opposite to the sun where the gegenschein appears is symmetrically situated with respect to the eastern and western light pyramids, that is, with respect to the eastern and western portions of the atmospheric 'lens' directly illuminated by the solar rays. It thus receives scattered light from both these sources and hence is more luminous than the surrounding regions which receive light from only one of the sources. The region therefore appears as a bright patch. Schmidt, however, does not explain why the centre of the gegenschein oval is not exactly at the anthelion.

It should also be remarked that Schmidt does not say anything about the nature of the particles constituting the scattering lens—whether they are air molecules or dust particles of some sort.

(iii) *Atmospheric theory* (*Hulburt and Vegard*).—Hulburt assumes that the exosphere or the 'fringe' region of the atmosphere is the scattering cloud of zodiacal light [see Chapter I, Sec. 5c]. Some of the high flying molecules in the spray region are ionized and those at levels beyond 30,000 km. will, under the combined solar radiation pressure and the earth's magnetic and gravitational fields, form a sort of oblong ring around the earth. On the daylight side of the earth, the ring lies roughly in the plane of the equator and on the night side it is warped off the equatorial plane approximately in the plane of the ecliptic, and is stretched out into a long oval by light pressure [23]. The ions in the ring are assumed to absorb the extreme ultraviolet light of solar radiation and re-emit as visible light a portion of the absorbed energy. This, according to Hulburt, is the zodiacal light.

The theory is capable of explaining most of the characteristic seasonal variations but leaves unexplained the similarity of the continuous part of the spectrum of the zodiacal light with that of the sun.

Another version of the formation of the atmospheric 'lens' is given by Vegard [24]. According to Vegard, the terrestrial atmosphere is topped by a 'corona', particularly on the daylight side. The process of formation of the corona is as follows: It is assumed that the sun emits radiation in the extreme ultraviolet, of wavelengths corresponding to those of soft X-rays. The atoms and molecules in the topmost layers of the atmosphere are ionized by these radiations and electrons of great energy are liberated. The ejected electrons as they move away—which they do more freely in an upward direction being less hindered by collision—produce an electric field between them and the positive ions left behind. This electric field not only prevents the photoelectrons from escaping from the terrestrial atmosphere, but also drives upwards to some extent the positive ions. Some of the neutral molecules are also carried upwards by collision with the ions. This region of electric double layer formed with electrons, ions and neutral particles, crowning the highest layers of the atmosphere, is called by Vegard terrestrial 'corona'. Due to the action of the terrestrial magnetic field the corona, consisting as it does of charged particles, bulges out near the equator and assumes a lenticular shape. The 'principal plane' of the lens coincides with the magnetic equator and the main extension of the bulge is on the daylight side. This bulge on the top of the atmosphere, illuminated by solar rays, appears as the zodiacal cone from the evening zones of the earth; further, since the extension of the bulge is towards the sun, the cone appears to lie in the plane of ecliptic.

No quantitative data are given by the author regarding the shape and size of the bulge, though it appears that its extension is much less than that of the 'fringe', being only a few hundreds of kilometres above the auroral layer at 70–100 km. height; the average space density of the particles is also much higher due to the presence of neutral molecules. It is interesting to recall that for the case of Hulburt's 'fringe', the ejected particles derive their energy by collision with particles in the denser atmosphere below. For the case of the 'corona', the electrons derive their energy from the high-energy photons absorbed by the molecules in the same region. Both the processes probably operate and contribute to the production of the 'fringe' region of the atmosphere.

(iv) *Planetary theory.*—As already mentioned the scattering cloud is assumed to belong to the sun in the planetary theory [6, 25]. It is supposed that the sun is surrounded by dust particles occupying a flat lens-shaped region extending well beyond the orbit of the earth in the plane of the ecliptic, and that the zodiacal band is the portion of the illuminated lens (by solar rays), as observed from the dark side of the earth. The particles cannot be of molecular size; they must be much larger, at least several wavelengths of light in diameter, as otherwise they would soon be dispersed by solar radiation pressure. The particles must also have planetary motion round the sun to prevent them from being sucked into it by gravitational attraction. To explain the zodiacal illumination of the

sky, the particles need not be numerous. If they are assumed to be 1 mm. in diameter having reflecting power of only 0·073 (the *albedo* of the moon), it can be shown that the observed intensity is accounted for if the average distance between the particles is 8 km.

According to Hoffmeister [26] the observations can be best explained by assuming the existence of two regions of maximum number-density of the particles round the sun, one within the earth's orbit and the other beyond Mars within the zone of minor planets.

An ingenious explanation of the gegenschein is given in the planet-dust theory by the application of a particular solution of the three-body problem in celestial mechanics to the case of the earth, the sun and the dust particle [27]. It can be shown that on the side of the earth opposite to the sun and on the line joining the two, there is a point of equilibrium round which a particle, assumed to be of infinitesimal mass (ideally), will move in a closed elliptical path if it entered the region near the point with appropriate initial velocity. If the particle be not of infinitesimal mass, it would move in a *nearly* closed orbit, and would remain in the region for a considerable time. Now, in the planet-dust theory it is supposed that a very large number of small particles are moving round the sun roughly in the plane of the ecliptic. It is, therefore, possible that some of the particles would pass near the equilibrium point with the correct initial velocity. Such particles would make one or more circuits around this point before pursuing their courses round the sun. There would, therefore, be a concentration of particles in this region and more light will be scattered from it than from the less populated surrounding region. There will thus be a glow of light in the dark anthelion region of the sky which is the so-called gegenschein. The effect will be enhanced because, as observed from the earth, the illuminated particles would all appear in full phase. The correctness of the theory, of course, depends on whether or not there are sufficient number of particles with the required initial motion to cause them to swarm round the equilibrium point in the manner described above.

New light has been thrown on the physical characteristics of the interplanetary dust cloud by the works of van de Hulst [29] and of Allen [30]. It appears that the so-called *F*-component of the continuous light of the outer solar corona (see Appendix, Sec. 8) is nothing but scattered light from the dust cloud [27*a*] and is thus merely an extension of the zodiacal light towards the sun. From an analysis of the photometric data of the corona and of the zodiacal light, Hulst has deduced the following characteristics of the interplanetary dust cloud. The thickness of the cloud perpendicular to the plane of the ecliptic is $\frac{1}{10}$th of the diameter of the earth's orbit and the total mass of the particles within the orbit of the earth is 5×10^{8} gms. (i.e. 10^{-9} times the mass of the earth). The space density is thus 5×10^{-21} gm./cm^3. The sizes of the particles range up to 0·3 mm. Particles with radii larger than 0·35 mm. are less abundant. The mean free path of the particles is 10^{19} cms., i.e. a million times the diameter of the earth's orbit. Hence the interplanetary dust cloud is extremely

transparent and any interaction between the scattering particles may be neglected.

Allen, who has assumed a non-uniform distribution of the space density of the particles (instead of a uniform density as assumed by Hulst), has come to similar conclusions regarding the nature of the scattering cloud. Incidentally, the distribution, in certain regions, as assumed by him agrees with the empirical distribution of Hoffmeister as mentioned above.

In regard to polarization, the planetary theory has this difficulty: Polarization is a characteristic of the so-called Rayleigh scattering by particles of molecular dimensions. How then, can the light scattered from such large-sized particles (0·3 mm. in diameter), as assumed in the planet dust theory, be polarized? The difficulty may be got over, if it is recalled that light reflected from minerals such as granite is partially polarized, and if it is assumed that the zodiacal particles are made of such minerals [31]. The atmospheric theory—so far as it assumes the scattering cloud to be an extension of the terrestrial atmosphere—has also its difficulty regarding polarization. It is well known that the intensity of light due to Rayleigh scattering varies inversely as the fourth power of the wavelength. If, therefore, the scattering particles—as assumed by Hulburt—be air molecules, then the intensity distribution in the zodiacal light spectrum ought to follow this law. No such distribution of intensity has, however, been observed in the zodiacal spectrum. It should, however, be remembered that the zodiacal light contains a large percentage of unpolarized light the origin of which may not be due to scattering. This unpolarized component may mask the inverse fourth-power distribution.

As regards the origin of the dust cloud it is sometimes suggested that the cloud particles are remnants of the matter sprayed out of the sun when the planets were born of the same due to the close approach of a passing star.

A more definite theory has been advanced by Fessenkov and is described in the sub-section to follow.

(v) *Planetary-cum-atmospheric theory* (*Fessenkov*).—According to Fessenkov both the planetary and the atmospheric theories have to be invoked to explain the whole group of phenomena—zodiacal pyramid, zodiacal belt, gegenschein, false zodiacal light and extension of the base of the upright zodiacal cone. The first two of these are ascribed to cloud of dust particles with the sun at its centre (planetary theory) and the rest to an extension of the terrestrial atmosphere, in the form of a tail, in the direction away from the sun (atmospheric theory). Both the dust cloud and the atmospheric extension are situated in the plane of the ecliptic being compressed perpendicular to the same.

Regarding the formation of the dust cloud Fessenkov advances the hypothesis [28] that the dust particles are produced by collision of sporadic meteors [see Chapter IIIB, Sec. 2] with asteroids. In support of his theory, Fessenkov quotes a parallel phenomenon, namely, the formation of thick clouds of dust on the surfaces of moon and Mercury and probably

on those of all members of solar system without atmosphere. Since the combined surface of all the asteroids is very large, the amount of dust particles produced is also large. Further, unlike the dust particles produced on the surfaces of the massive heavenly bodies, the dust particles produced on the surfaces of asteroids cannot be retained by them as the attraction is very small. According to Fessenkov, they are drawn into the solar system from different directions and with different velocities and form an oblate spheroid with the sun at its centre. The spheroid is surrounded by a thick ring of dust in the asteroid zone. This spheroid, according to Fessenkov, gives rise to the conical zodiacal light and the ring of dust surrounding it the zodiacal band along the ecliptic.

To explain the gegenschein, the false zodiacal light and the broadening of the base of the zodiacal cone (when its axis makes small angle with the vertical) Fessenkov suggests an atmospheric theory. According to Fessenkov the outer and the most extended portion of the terrestrial atmosphere has a shape as shown roughly in Fig. 10. In the direction of the sun, this outer atmospheric layer may be represented by the surface of an elliptic paraboloid extending one or two thousand kilometres and flattened in the direction perpendicular to the plane of the ecliptic. In the direction opposite to that of the sun it is confined almost entirely in the plane of the ecliptic and widens in the shape of a cone with an angle of about 8°–10° between the generatrices. The density of the matter in the cone decreases slowly being about halved for each 4·7 earth radii. Fessenkov has made detailed study of the luminous effects which this tail would produce in the sky when viewed from the night side of the earth and has shown that they can explain, at least qualitatively, the gegenschein and some zodiacal light phenomena not explained by the dust-cloud theory. For example, the shape and size of the gegenschein oval, its nightly parallactic motion and its opening out to form the false zodiacal light, all conform to results as expected from the hypothesis. Further, the fact that the continuous spectrum of the zodiacal light resembles that of the solar spectrum, but that the continuous spectrum of the gegenschein [21*c*] is identical with that of the night air-glow, shows that the origin of the two phenomena must be different. The former is due to dust particles (as assumed in the planetary theory) and the latter to extension of the terrestrial atmosphere (as assumed in the atmospheric theories of Vegard, Hulburt, and others). However, the broadening and the increased luminosity at the base of the zodiacal light cone is to be ascribed to contribution of luminosity from the tail of the terrestrial atmosphere (as evidenced by the spectral study). Fessenkov thus remarks [28*a*]: 'The old controversy

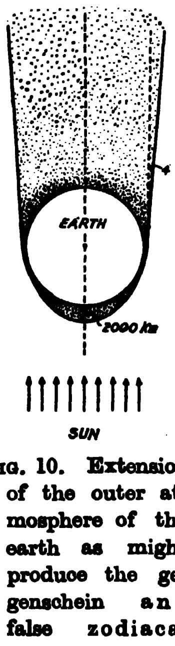


FIG. 10. Extension of the outer atmosphere of the earth as might produce the gegenschein and false zodiacal light according to Fessenkov.

concerning the nature of the zodiacal light is ended; the phenomenon is partly of a cosmic and partly of an atmospheric nature. It is certain that the basic part of the phenomena is connected with the scattering of light by meteoric matter in the interplanetary space; but it also depends, in part, on the terrestrial atmosphere. The upper layers of the terrestrial atmosphere cannot therefore be symmetrical with respect to the earth's surface.' (In this connection, one may recall Hulbert's hypothesis (*vide* sub-sec. iv) in which it is assumed that on the night side, the atmosphere is stretched into a long oval by light pressure and is warped off the equatorial plane approximately in the plane of the ecliptic.)

3. GALACTIC LIGHT

When the effects of the star light, the zodiacal light, the light of the air-glow and the scattered light due to all these have been taken into account, there still remains some excess of illumination. This light is supposed to be due to light from galactic nebulae and light from the milky-way scattered by interstellar matter. It is now well established that the space between the solar system and the stars is not entirely devoid of matter. It is found

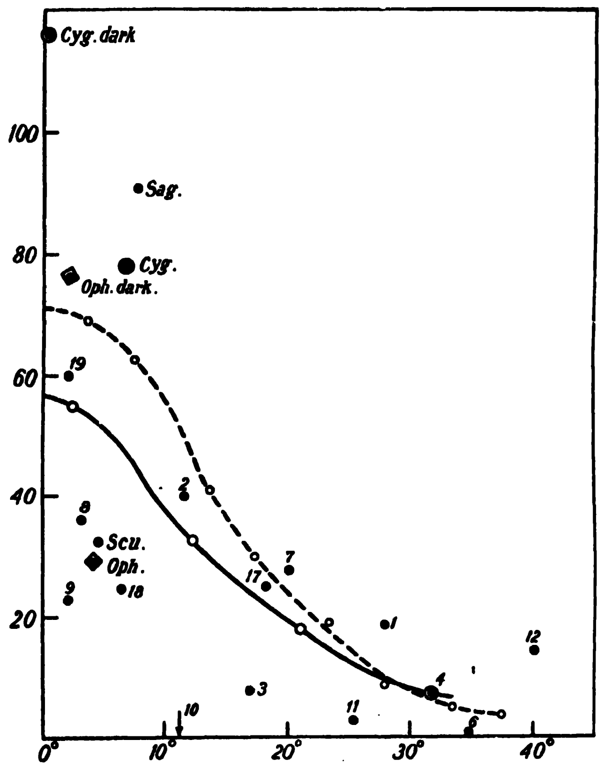


FIG. 11. Variation of mean galactic light with galactic latitude. The dashed curve is for all the data. The continuous curve represents data for the same regions for which mean star counts have been made. (After Elvey and Roach.)

to contain atoms of hydrogen, calcium, sodium, potassium, iron, oxygen and titanium—some of them ionized, also molecules of CN and CH [33]. The space density of matter is very small being of the order of one particle per $cm.^3$

Fig. 11 depicts the variation of galactic light with galactic latitude [13]. The ordinates represent brightness due to galactic light in units of tenth photographic magnitude per square degree of the sky. It will be seen that, as expected, the maximum brightness of the night sky due to this source lies at zero galactic latitude; at this latitude the brightness corresponds to 57 stars of tenth photographic magnitude per square degree.

4. NIGHT AIR-GLOW—SPECTROSCOPIC STUDIES

(a) Introduction

As mentioned in Sec. 1 an important proportion of the luminosity of the night sky is due to the self-luminesence of the upper atmospheric gases, or, to what has been described as 'air-glow'. The study of the night air-glow is of great importance as it helps to understand many of the physical properties of the upper atmospheric regions. In what follows we shall first give a brief account of the results that have been obtained from spectroscopic and photometric studies of the air-glow. The origin of the air-glow, the modes of excitation of its spectral components and the possible causes of the variations of their intensities (nightly, annual and with solar cycle) will then be discussed. It will be seen that besides ozone, atomic oxygen and atomic nitrogen (to which references have already been made in Chapters IV and V) the upper atmospheric regions also contain hydrogen and sodium.

(b) The Spectrographs

The night air-glow being very faint the spectrograph used must be of high light-gathering power. Typical dimensions of the lens and the prism systems used by some of the earlier and more recent workers are given below. It will be noted that with the design of the apparatus as used by the earlier workers the exposure necessary had to be from several hours to some tens of hours.

Workers in India have used spectrograph with a single flint prism and with collimator of focal length 12 inches and diameter 1·5 inch [16]. The camera had a Dallmeyer lens of focal length 2 inches and aperture $f/1{\cdot}5$. The length of the spectrum between the oxygen green line and the H-line of the solar spectrum was about 3·5 mm. For obtaining a good spectrum an exposure of about 26 hours (spread over 5 nights) was necessary.

Dufay employed a quartz spectrograph for the visible and the near ultraviolet regions [34]. The collimator and the objective were of 9 cm,

focal length and 25 mm. diameter, the aperture being $f/3 \cdot 5$. The prism was a 60° one, composed of two similar right-angled quartz prisms, one right-handed and the other left-handed, cemented together. The dispersion was 120 A per mm. at λ 3500. With slit width 0·2 mm., spectrograms could be obtained with 11 hours exposure. With finer slits, total exposure of 48 to 60 hours was necessary.

For specially studying the region towards the red end of the spectrum Cabannes used a spectrograph with a flint prism of 115 mm. height with a 205 mm. base [35]. The camera objective had a focal length of 80 mm. The aperture of the apparatus was large, being $f/0 \cdot 07$. The spectral range $\lambda\lambda$ 5265–7280 occupied a length of 4·6 mm. With a slit of 0·2 mm. width, exposures varying from 6 to 17 hours were necessary for securing good photographs on clear nights.

Lord Rayleigh used a spectrograph of large aperture $f/0 \cdot 09$, the focal length of the objective being only 30 mm. A flint prism was used.

Elvey and Farnsworth have used a spectrograph with collimator 20 inches and camera lens 2 inches in diameter with aperture $f/0 \cdot 66$ [36]. The dispersing system was 'one of the Mantois prisms of the Bruce spectrograph of the Yerkes Observatory'. The linear dispersion of the apparatus was 1500 A per mm. at the sodium *D*-lines. For photographing the spectrum during twilight (instrument pointing 5° above the horizon in the direction of the setting or the rising sun) an exposure of 10 minutes sufficed to produce a measurable density of the spectrogram. During dark hours of the night, the standard exposure for a zenith distance of 70° was 90 minutes.

For studying the infra-red region 6000 A to 8800 A, Meinel used a spectrograph consisting of a replica grating and a flat-field Schmidt type camera of an effective speed of $F/1$ [37]. The grating size was 100 mm. by 165 mm. with 7500 lines per inch and maximum energy concentration (about 80 per cent) in the first order at about 8000 A. The camera had a dispersion of 250 A/mm. in the first order and a focal length of 14 cm. The slit is placed at a large distance from the grating; as such, there is no need of a collimator. Hypersensitized Eastman 1N emulsion was used. With projected slit width 50μ and 30μ the exposures were 7 hrs. and 32 hrs. respectively. In another design with dispersion of 247 A/mm. in the first order and a photographic resolution of approximately 10μ excellent spectra were obtained with 12 hrs. exposure with projected slit width of 21μ [38].

It is hardly possible to use photographic plates for study in regions beyond 9000 A on account of impossibly large exposures required. Photoelectric pick-up may be used instead. But this has the disadvantage that on account of the wide spectral bands of sensitivity of photoelectric surfaces, the spectral resolution is very small. The ideal solution is, of course, to use monochromator with photoelectric pick-up. An approach to this may be made by the use of interference type of filters, which transmit only narrow bands of wavelength, in conjunction with photoelectric photometer. Such an apparatus will be described in Sec. 5*b*.

(c) The Spectrum

(i) *General description.*—The night air-glow spectrum consists of a system of lines and bands on the background of a continuous spectrum.

Continuous spectrum.—The continuous spectrum has been found to be crossed by absorption lines coinciding with some of the Fraunhofer lines. The intensity of the spectrum towards the blue and violet end is much stronger than what one would expect from pure scattering. It is known that the intensity of sunlight scattered by the atmosphere decreases rapidly towards the short wavelength side from the violet end. A photometric comparison of the distribution in the ultraviolet part of the continuous spectrum of the night air-glow with that of the scattered skylight during daytime had been made by Dobronravine and Khvostikov [39] during the years 1937 and 1938 at the astronomical observatory at Siméise, Crimea. Fig. 12 depicts the results of an observation. It will be seen that the night air-glow spectrum is relatively richer in ultraviolet than the scattered light

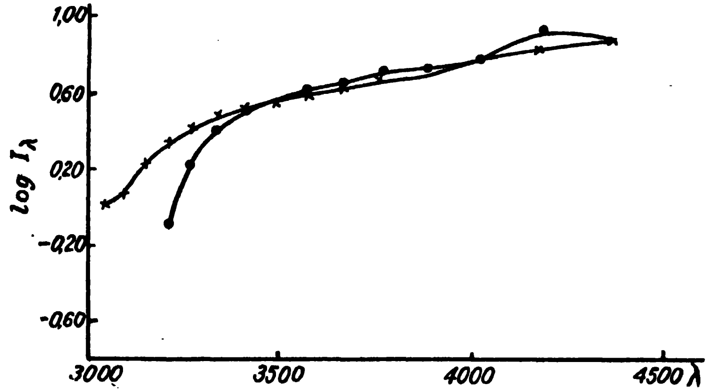


FIG. 12. Comparison of the variation of intensity with wavelength of scattered sky light (during daytime) with that of the continuous spectrum of night air-glow. (—•— Scattered skylight; —×— night air-glow.) It will be noticed that the night air-glow spectrum is richer in ultraviolet. (After Dobronravine and Khvostikov.)

from the day sky. No clear explanation of this curious fact is known. It has been suggested that a major part (more than 50%) of the continuum is due to superposition of wings and weak bands of the Vegard-Kaplan system [39*a*]. Other lesser contributions are by zodiacal light and by unresolved stars. Suggestion has also been made that the continuous spectrum is emitted in the process of radiative association of neutral atoms [39*b*].

It has been suggested that night air-glows of abnormal brilliance are due to enhancement of the continuous spectrum radiation originating in the upper atmosphere and not to extra-terrestrial sources like galactic or zodiacal light. According to Barbier [40] about 2/3 of the background intensity in the blue, violet and ultraviolet is to be attributed to the 'wings' of the Vegard-Kaplan and Herzberg bands.

Blackett [41] has discussed the interesting possibility of contribution to the continuous part of the night air-glow spectrum by the so-called Cerenkov radiation emitted by cosmic rays. (When a transparent material is traversed by fast electrons, a continuous spectrum of visible light, mainly in the direction of motion of particles, is emitted. This was discovered by Cerenkov.) Blackett points out that if the particle velocity is very nearly that of light then the same type of radiation should be produced in a gas. According to his estimate the intensity of the Cerenkov radiation at sea-level, due to bombardment of the atmosphere by cosmic rays (whose velocities approach that of light), is about 10^{-4} of the total light of the night air-glow. The contribution is thus negligible for clear night sky. However, the radiation should exist even under an opaque layer of very high clouds. 'The Cerenkov radiation of cosmic rays sets a limit, even if an unrealizable one, to the darkness of the darkest night.'

Line and band spectra.—In Table I the observed lines and bands with their origins are listed. It will be noticed that the most prominent lines and bands in the visible and near ultraviolet are those due to atomic oxygen, atomic sodium and molecular nitrogen. (It will be seen in Sec. 9 that though the sodium *D*-lines are very prominent, sodium atoms constitute a very small fraction of the atmosphere.) The near infra-red region is dominated by the rotation-vibration bands of the OH radical. In the ultraviolet the strongest bands are due to the Herzberg system of O_2.

The list as in Table I does not include many lines and bands that are still unidentified.

In Figs. 13 and 14 (Plate I) spectrograms of night air-glow are reproduced. The spectrum in Fig. 13 is for the visible and near ultraviolet region. It was taken by Elvey with the nebular spectrograph of the McDonald Observatory, Texas, U.S.A. [42]. In Fig. 14 two types of spectra for the red and near infra-red region taken at the Lick Observatory, U.S.A., are shown [43]. *A* is for a quiet night; *B* is for a night when a faint aurora was present in the north. The spectra were taken by Meinel with transmission replica grating. In the quiet night sky spectrum (*A*) a large number of rotation-vibration bands of OH in the ground ($X\ ^2\Pi$) state can be recognized by comparing with Fig. 16 (Plate II). The red atomic oxygen lines at λ 6300 and λ 6363 appear clearly. The atmospheric *A*-band system of O_2 at λ 8628 and λ 8659 appear in emission. The broad absorption features at λ 6900, λ 7200, λ 7600 and λ 8220 correspond respectively to the Fraunhofer *B*, *a*, *A* and *Z* bands. The special features in the spectrum when a faint trace of aurora was present (*B*) are the following: A group of very intense radiations near λ 7700 which is identified with (2, 0) transition of the first positive bands of N_2. Another group at about λ 8600 also include first positive system of N_2 transitions (2, 1). The most striking feature is, however, the permitted lines of atomic oxygen (λ 7774 and λ 8446) and of atomic nitrogen (λ 8684 and λ 8630). (See Appendix, Sec. 2, Fig. 16.) (It is to be noted that 'as a result of widening, the reproduction shows many spurious faint lines which are

TABLE I

Lines and bands in the night air-glow spectrum

Origin	Transition	Remarks
O atom	$^1S \rightarrow {}^1D$ (λ 5577)	Green oxygen line. Permanently present in the night air-glow spectrum. Strong intensity. Transition forbidden.
	$^1D \rightarrow {}^3P$ (λ 6300) (λ 6363) (λ 6392)	Red oxygen lines. Permanently present, but of lesser intensity than λ 5577. Transition forbidden. Considerably enhanced in twilight spectrum.
N_2^+ ion	$^2\Sigma_u \rightarrow {}^2\Sigma_g$ ($0 \rightarrow 0$ λ 3914)	First negative bands of N_2^+. Absent or very faint in the night air-glow; but observed when the slightest trace of aurora is present and also in the twilight.
N_2 molecule ..	$B^3\Pi_g \rightarrow A^3\Sigma_u$ (red and infra-red region)	First positive group of N_2; fairly strong. Always present.
	$A^3\Sigma_u \rightarrow X^1\Sigma_g$ (blue-violet region)	Vegard-Kaplan bands. Strong and always present. Forbidden transitions.
	$a^1\Pi_u \rightarrow X^1\Sigma_g$ (ultraviolet region)	Lyman-Birge-Hopfield systems. Many coincidences with the computed band heads of the system show that this system is most probably present.
Na atom	$^2P \rightarrow {}^2S$ (λ 5894)	Sodium *D*-line. Strong; always present. Greatly enhanced in twilight spectrum.
O_2 molecule ..	$B\,{}^3\Sigma_u \rightarrow X\,{}^3\Sigma_g$ (blue and ultraviolet region)	Runge-Schumann bands. There is a fair probability of the occurrence of this band system.
	$A\,{}^1\Sigma \rightarrow X\,{}^3\Sigma$ (infra-red region)	Atmospheric band system. Identified by Meinel
	$^3\Sigma_u^+ \rightarrow {}^3\Sigma_g^-$ (ultraviolet region)	Herzberg bands. Identification almost certain.
OH radical	Vibration-rotation bands of the ground state $X\,{}^2\Pi$ (red and near infra-red).	Meinel bands. Very strong.

due solely to random fluctuations in the silver grain distribution over the width of the spectrum on the original plate.')

Critical discussions of the night air-glow spectra are to be found in references [44, 45, 46, 47, 48, 49, 49*a*]. In what follows the characteristics of the observed lines and bands will be briefly discussed.

(ii) *Atomic oxygen lines.*—The atomic oxygen lines are λ 5577, λ 6300 and λ 6363. They are all due to the so-called forbidden transitions (see Fig. 17). The line λ 5577, well known as the *green auroral line*, is the most important line and comes out with great intensity in the spectra of both the aurora and the night air-glow. The presence of this line was for a long time known in the aurora though its origin was unknown till 1924. It will be discussed further in Section 4*d*.

With even a faint trace of aurora allowed lines at λ 7774 and λ 8446 appear. (See Appendix, Sec. 2, Fig. 16.)

(iii) N_2 and N_2^+ *bands.*—The most important bands of N_2 are the Vegard-Kaplan bands (groups near λ 4420 and λ 4200). They were first noticed by Rayleigh as early as 1922 due to their large intensity. Their origin, however, was for a long time uncertain but they are now known to be due to the forbidden transitions $A^3\Sigma_u \rightarrow X^1\Sigma_g$ [see Appendix, Sec. 2, Fig. 17] of nitrogen molecules and the story of their identification is interesting. Kaplan who had been studying for a long time the mode of excitation of the night air-glow and auroral lines in the laboratory, identified a number of bands as belonging to certain inter-combination bands of N_2 [50]. Previous to this, Vegard had observed these bands while studying the effect of bombardment of a mixture of solidified argon and nitrogen dioxide by cathode rays [51]. Their classification was, however, left to Herzberg [52]. Kaplan identified a large number of lines of these bands with those occurring in the spectrograms obtained by Cabannes and Dufay and definitely established that a number of groups of lines belongs to this band system. Subsequently Cabannes and Dufay identified about 32 lines of the night air-glow in the region $\lambda\lambda$ 3500–5000 as belonging to this system [53]. The following are the strongest bands of the system (1, 11) λ 3707, (6, 15) λ 3834, (2, 13) λ 4702, (3, 14) λ 4174, (2, 14) λ 4420, (3, 15) λ 4543 and (2, 15) λ 4827.

The first positive group of nitrogen bands is also quite strong. They extend from the green to the infra-red region. The lines from λ 5320 to λ 5510 are due to transitions $v'-v'' = 5$; from λ 6040 to λ 6170 to $v'-v'' = 4$; from λ 6350 to λ 6780 to $v'-v'' = 3$; from λ 7250 to λ 7480 to $v'-v'' = 2$ and from λ 7750 to λ 8330 to $v'-v'' = 1$ [36].

A number of coincidences with the band-heads of the Lyman Birge-Hopfield system (see Appendix, Sec. 2, Fig. 17) have been observed. It is likely that the band system is present in the night air-glow spectrum [53*a*, 53*b*].

The first negative bands due to N_2^+ are generally absent from the night air-glow though they appear with great intensity in auroras. Some workers maintain that the bands may be observed in the night air-glow even when the sky is free from all auroral displays [54]. This has, however, been questioned by other investigators. These bands may be obtained in the spectrograms of the evening and of the morning sky when the last rays of the sun are disappearing, or, when the first rays are illuminating the

upper atmosphere [55] (see Fig. 29, Plate II). Reference to this interesting twilight effect will be made later.

(iv) *Atomic nitrogen lines.*—Atomic nitrogen lines (both permitted and forbidden transitions) are present when there is even a faint trace of aurora. (See Chap. VIII, Sec. 3*a*; Fig. 15. See also Appendix, Sec. 2, Fig. 15.) But their presence in the quiet night air-glow spectrum is doubtful. This, however, does not exclude the possibility of the presence of nitrogen atoms in the upper atmospheric regions. (See Chapter V, Sec. 3.)

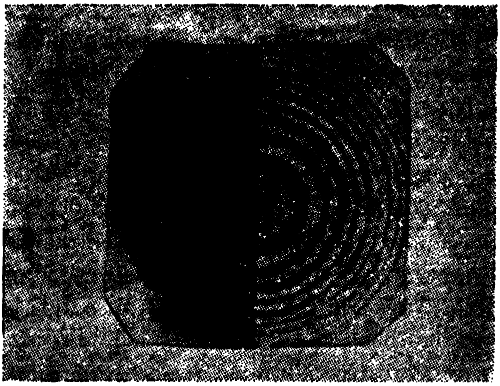


FIG. 15. Comparison of interference fringes due to yellow radiation in the night sky (Montpellier, France) with the sodium *D*-line (0·15 mm. etalon).

(v) *Sodium D-lines.*—The presence of the sodium *D*-lines in the night air-glow spectrum is proved conclusively by wavelength measurements by Fabry-Perot interference rings [56, 57]. (See Fig. 15.)

(vi) O_2 *bands.*—The band systems due to O_2 lie either in the infra-red or in the ultraviolet or in a region where bands due to N_2 predominate. (See Appendix, Sec. 2, Fig. 19.) It is therefore difficult to identify them with certainty. Nevertheless, identification of the Herzberg bands ($^3\Sigma_u^+ \longleftrightarrow {}^3\Sigma_g^-$) in the ultraviolet appears to be well established.

The higher members of the so-called atmospheric *A*-band system ($A^1\Sigma \rightarrow X^3\Sigma$) have also been definitely identified [43]. (It may be noted that these bands were first observed by Kaplan in the afterglow in pure O_2. *Vide infra.*) In Fig. 14 (Plate I) the emissions at λ 8628 and λ 8659 coincide with two maxima of the (0, 1) transition [58]. There are also several other coincidences each with lower level above the zero vibrational level. Hence, as expected, none of these emissions show definite enhancement during twilight. The intensity of the bands in the night air-glow spectrum is extremely variable.

A broad absorption feature at λ 7600 may also be noticed in the same figure (Fig. 14). This corresponds to the (0, 0) atmospheric band of O_2. This shows that bands of O_2 which would otherwise be expected strong in emission (in view of the presence of the (0, 1) band) are absorbed by the underlying amount of gas. (Other broad absorption features are also to be

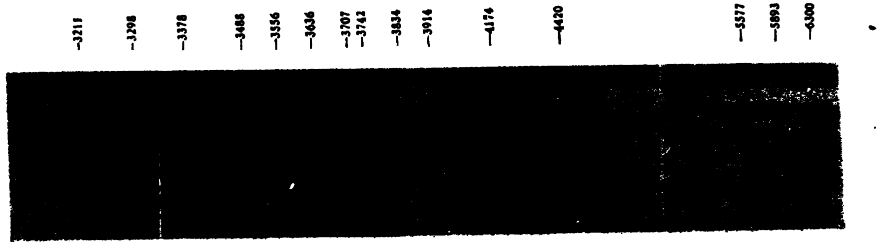


FIG. 13. Spectrogram of night air-glow taken by Elvey with nebular spectrograph of the McDonald Observatory, Texas, U.S.A.

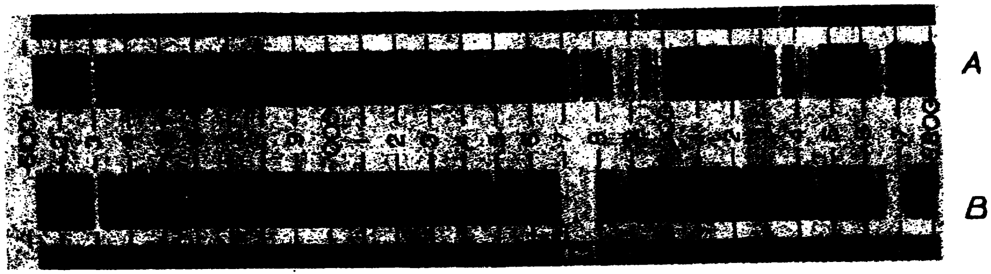


FIG. 14. Two types of night air-glow spectra for the red and near infra-red regions taken by Meinel at the Lick Observatory, U.S.A., with a transmission replica grating. *A* is for quiet night sky; *B* is for a night when a faint aurora was present in the north.

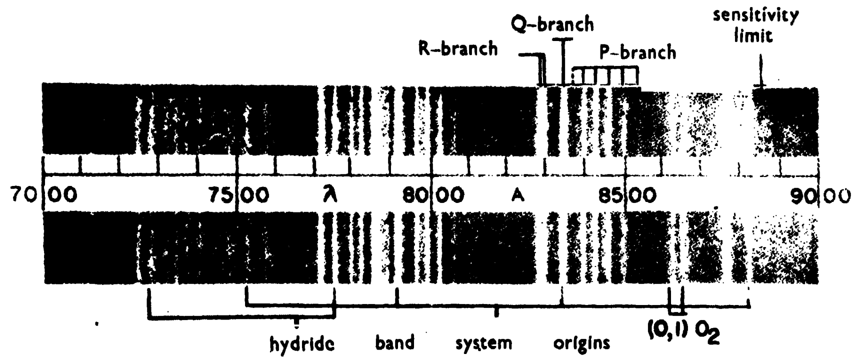


FIG. 16. Infra-red spectrum of the night air-glow illustrating the OH-bands. The spectrum was taken by Meinel at the Yerkes Observatory with transmission replica grating.

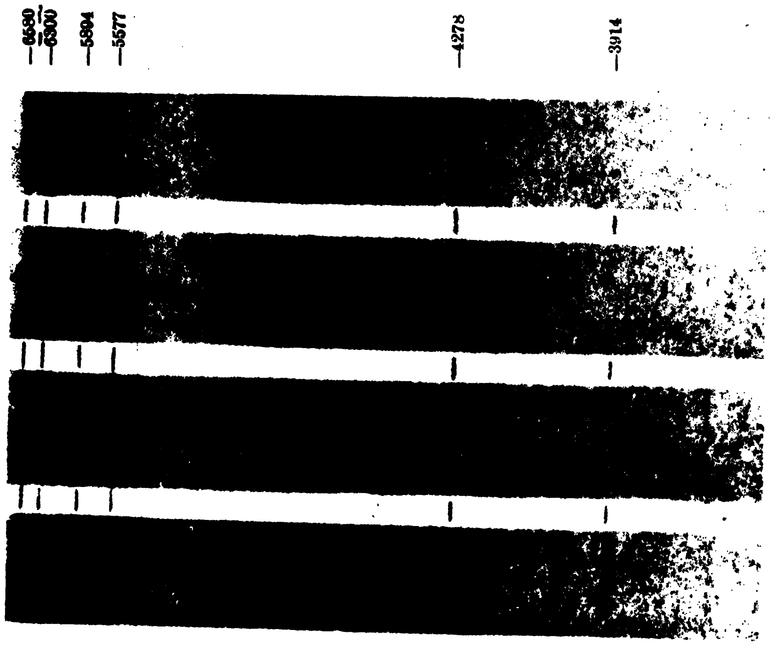


FIG. 29. Spectrograms of the night sky taken with a nebular spectrograph at the McDonald Observatory, Texas, U.S.A., illustrating the enhancement of the red oxygen lines, the appearance of the negative band system of ionized nitrogen molecule and the increase of the intensity of the sodium line with the approach of twilight. The apparatus was pointed towards the eastern horizon; the exposures from bottom to top are three hours and fifty minutes, one hour, half an hour and fifteen minutes respectively. The topmost exposure was given during dawn and the one below just preceding the astronomical dawn.

noted at λλ 6900, 7200 and 8220. They correspond respectively to the Fraunhofer *B*, *a* and *Z* band of H_2O.)

Many coincidences with the Runge-Schumann band systems ($B^3\Sigma \rightarrow X^3\Sigma$) have also been reported. However, the identifications cannot yet be considered as quite satisfactory.

(vii) *Meinel bands* (*OH*).—The important identification of bands due to the OH radical in the night air-glow spectrum in the red and near infra-red regions is due to Meinel [59]. These are rotation-vibration bands ($v \leq 9$) of the ground state $X^2\Pi$. The identifications have removed the long standing puzzle about the origin of the strong emissions observed at λ 6560 and probaby also at λ 10,400 [60]. The former is the unresolved *P*-branch of the (6, 1) band and the latter a blend of the (9, 5), (4, 1) and (5, 2) bands. (See Appendix, Sec. 2(*b*), Fig. 8.) In the spectrogram in Fig. 16 (Plate II) the emissions of these bands in the region λλ 7000–9000 are shown. The spectrogram was taken at the Yerkes Observatory, U.S.A., with transmission replica grating. The exposure was of 12 hours on hypersensitized Eastman 1N emulsion with a projected slit width of 21 microns (see Sec. 5). The spectrum in the range λλ 8000-11000 has also been studied by Krassovsky [60*a*] using a prism-spectrograph and a Cs–O–Ag electron image converter. The emissions recorded by Krassovsky in this range correspond to those of the Meinel bands. A strong emission was found at 9976 A. This is believed to be due to a blend of the strong (9, 5) OH band at 10014 A and of the (0, 2) O_2 band.

It should be mentioned that the identification of the OH bands as due to the rotation-vibration bands of OH is due to Gerhard Herzberg (see p. 564, ref. [59]).

Other band systems.—Coincidences with bands due to H_2O, NO, CO, O_2, NH, CH, CN have been reported from time to time. The identifications are, however, either accidental or only tentative. Regarding CO a new system as suggested by Barbier seems probable [61].

(*d*) Laboratory attempts to produce the air-glow spectrum

(i) *Introduction.*—Attempts have been made to reproduce in the laboratory the spectrum of the air-glow and also of the aurora [see Chapter VIII, Sec. 4]. Such investigations are helpful for identifying lines or bands of unknown or doubtful origin and also in indicating the probable conditions prevailing in and the probable modes of excitation in the upper atmosphere. It should, however, be borne in mind that the laboratory conditions of excitation are necessarily widely different from those in the upper atmosphere. For instance, in the upper atmosphere there is no 'wall effect' as there must be in a discharge tube. Further, in the case of air-glow the ultimate source of its energy is the ultraviolet solar radiation, while that for the glow in the discharge tube is the fast electrified particles which have acquired momentum in the field of the exciting voltage. It is therefore hazardous to draw any conclusion regarding actual processes of excitation of the air-glow or auroral spectra, from considerations of the special condition

of pressure, gas mixture or wall condition under which the glow discharge is obtained.

The difficulty of reproducing the night air-glow (or the auroral) spectrum in the laboratory lies in the fact that most of the lines and bands of same are due to the so-called forbidden transitions. Such lines and bands can only be produced in the laboratory (for reason to be discussed presently) by using specially prepared discharge tubes and employing special modes of excitation. We give below short descriptions of methods by which it has been possible to produce the prominent lines and bands in the laboratory.

The first laboratory attempt at imitating auroral spectrum was made by Vegard by bombarding solidified nitrogen by cathode rays [51]. The yellow-green band thus obtained contained, as was found later, the Vegard-Kaplan bands in addition to the then known first positive bands of nitrogen; the atomic oxygen line λ 5577 was, however, not found. In recent years most of the important work on laboratory reproduction of the auroral and the night sky spectrum has been done by Kaplan. Before describing them we shall give an account of the classical work of McLennan and his associates on the excitation of the oxygen green line in a discharge tube.

(ii) *McLennan's work: Green atomic oxygen line.*—The origin of this most important line of the aurora and of the night air-glow was for a long time a puzzle to the spectroscopists. Attempts had been made to trace the origin of this line to solidified nitrogen [51], to Krypton, and to a hypothetical element Geocoronium. All such speculations were, however, set at rest when McLennan and Shrum [49] proved that the line is due to atomic oxygen. They studied the discharge in oxygen mixed with helium or neon, and found that in the spectrum of the discharge the green line λ 5577 could be identified easily. Further, this line could also be excited in pure oxygen under high current density. They concluded that the line must therefore be attributed to 'some hitherto unknown spectrum of oxygen'. McLennan, McLeod and McQuarrie [63] repeated the experiment and found that the intensity of the green line increased enormously when mixed with argon.

Since the mode of excitation of this important line is interesting from many points of view, we shall describe it in a little detail. In spectroscopic nomenclature the line is said to be due to the transition of the oxygen atom from ^1S to ^1D state [see Appendix, Sec. 2(*d*)]. Both are metastable states of the oxygen atom; an oxygen atom, if excited to ^1S state will remain there for an appreciable time (0·5 sec.) before spontaneous transition to a lower state (see Fig. 17). Now an excited atom in a discharge tube may be removed from its state of excitation by various processes. Firstly, it may radiate spontaneously and drop to a state of lower energy. Secondly, it might react with incident radiation and thus be raised from its original state to a state of still higher excitation or drop to a state of lower excitation. Thirdly, it might collide with an electron and give up its excitation energy to the latter as kinetic energy. Finally, it can collide with an atom or a

molecule either in the contained gas or on the walls of the discharge tube and be relieved of its energy by exciting the latter. The last-named mode

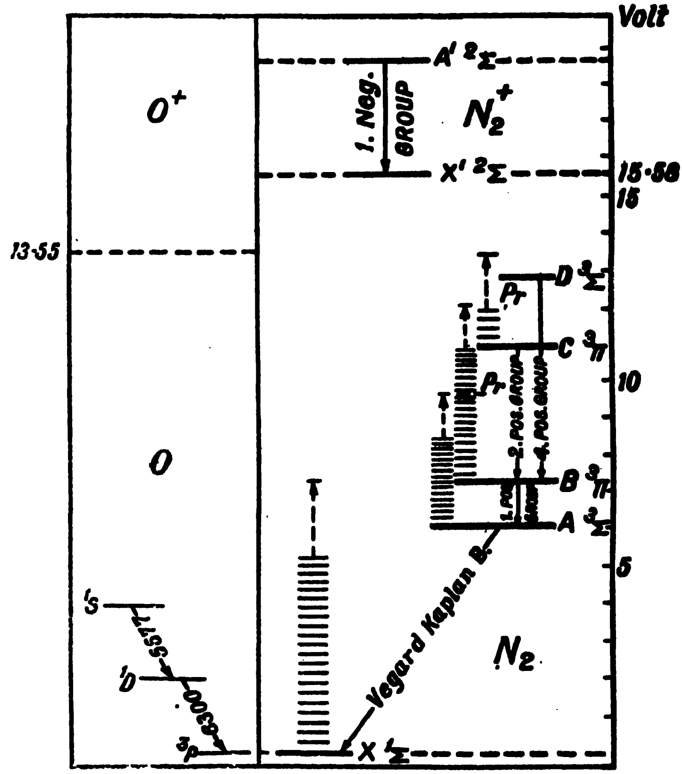


FIG. 17. Energy level diagrams of O and N_2 illustrating the emission processes in night air-glow spectrum. See also Figs. 16 and 17 in Appendix, Sec. 2, where the spectroscopic symbols are explained.

of transference of energy has a very high probability when the particle encountered possesses a state of excitation whose energy is about the same as that of the excited atom. A molecule, since it possesses a large number of energy states, is particularly effective in absorbing the energy of the colliding excited atom, as it is possible that one of its many states of excitation closely matches that of the excited atom. For the same reason collision with a solid which has a complicated structure and possesses innumerable excited states is very much effective in de-exciting the colliding atom. (It may be noted that the probability of a collision with an atom or a molecule in which energy of the excited atom is transferred to kinetic energy is very small.)

Now, as already mentioned, the oxygen atom in the metastable ^1S state has a life-time of about 0·5 sec. The probability of its removal from the excited state by spontaneous transition to a lower state, accompanied by radiation, is thus small. (Compare the life-time of ordinary excited atom $10^{-9}-10^{-8}$ sec.) In fact, in the ordinary discharge tube, this

probability is in general very small compared to the probability of its de-excitation by the various collisional processes described above, in particular by the collisions with the wall. Now it may be possible to minimize the chance of de-excitation by wall-collision by increasing the gas pressure, as this will check the diffusion of the excited atoms towards the walls. But the increased pressure will increase the chance of de-excitation by collisions with electrons and other molecules. There is thus an optimum pressure which is favourable for radiation from metastable atoms or molecules. In the experiments of McLennan, McLeod and McQuarrie [63] with pure oxygen a pressure of 2 mm. was found optimum for production of λ 5577 line. The highest intensity of the line was, however, obtained when the oxygen, only in very small quantity, was mixed with some noble gas like helium or neon [62] or argon [63]. The explanation is obvious. Since these gases have no excitation potential below 10–20 eV, they cannot take up, on collision, the energy from the oxygen atom in ^1S metastable state, the value of which is only 4·2 eV. The presence of these gases even at comparatively high pressure therefore, while not increasing the probability of transfer of the energy of excitation of the atoms by collision, prevents their diffusion to the walls of the discharge tube where they would otherwise have been de-excited.

We now understand why the λ 5577 line, as also the other lines and bands due to transitions from metastable states, are observable in the upper atmosphere. Here the pressure is low and, as a consequence, collisions which may lead to de-excitation are few; further there are no walls, the collisions with which in a discharge tube are so effective in removing the atoms from their excited states. An atom therefore excited to a metastable state has the chance of remaining in that state until it jumps spontaneously to a state of lower energy with the emission of the forbidden line.

(iii) *Kaplan afterglows.*—Kaplan has carried out extensive series of experiments to reproduce auroral and night air-glow spectra in the afterglows of discharges in pure nitrogen or in nitrogen mixed with traces of other gases (particularly oxygen) contained in specially 'conditioned' tubes.

It is found that under certain special condition of the walls (the nature of which is still ill understood) the glow in a discharge tube may continue for an appreciable time even after the discharge current has been cut off. The best known example of this is the nitrogen afterglow, also called *Lewis-Rayleigh* glow. Under favourable conditions the glow may persist for some hours. The spectrum of the glow consists of selected bands in the first positive system of nitrogen. The glowing gas is capable of reacting vigorously with many substances and hence is called 'Active Nitrogen' after Rayleigh. A full account of the nitrogen afterglow and properties of active nitrogen is to be found in ref. [64].

Many hypotheses have been put forward regarding the nature of the active substance which produces the afterglow phenomenon. A recent one put forward by Mitra [64] is as follows. According to this theory, the active substance is the N_2^+ (X') ions produced by the discharge. They persist because the walls of the discharge tube are so conditioned that recombination of the ions with electrons on the surface is prevented. This is believed to be due to the formation of an adsorbed

layer of N_2 (or, N) on the glass surface. Since N_2 molecules (or, N atoms) possess very little electron affinity, a layer of N_2 (or, N) on the surface prevents the formation of a negative surface charge by deposition of electrons. The N_2^+ ions therefore instead of being neutralized on the surface are neutralized in the volume by three body collisions as follows,

$$N_2^+ (X') + e + N_2 \rightarrow N_2(B\text{-state}) + N_2(A\text{-state}).$$

The ionization potential of N_2 is 15·58 eV and this amount of energy is released on its neutralization. The whole of this energy can be shared by the two neutral N_2 molecules produced if one is raised to a suitable vibrational level of the A-state and the other to that of the B-state (see Appendix, Sec. 2, Fig. 17). The probability of the reaction is thus high because the whole of the released energy is taken up as potential energy by the reaction products. If the molecule in the A-state be in its lowest level $v' = 0$ (6·1 eV) then the other molecule will take up the rest of the energy, namely 9·48 eV by being raised to vibrational level $v' = 12$ (9·75 eV) of the B-state. The emission of the first positive bands which is due to the transition $B \rightarrow A$ is thus explained. (From the A-state Vegard-Kaplan bands are emitted by the transitions $A \rightarrow X$. However, as the A-state is a highly metastable state the N_2 molecules in this state are de-excited by collision by a radiationless process.)

The Kaplan afterglows are produced in sealed discharge tubes which have been cleaned up by prolonged discharge (uncondensed) to get rid of impurities. These are generally oxygen from gas cylinder and hydrocarbons derived from stop-cock grease and they produce spectra of O, NO and CN. The prolonged discharge also 'conditions' the walls of the tube by formation of adsorbed layer. Kaplan remarks that in such tubes the catalytic effect of the wall has, in some way, been removed and the tube behaves effectively as if it has no walls. (Of the numerous papers published by Kaplan on the subject mention may be made of those in references [65, 66, 67, 68].)

Using pure O_2, Kaplan [72] found that the afterglow spectrum consisted solely of the forbidden $^1\Sigma \rightarrow {}^3\Sigma$ system of bands due to transitions from the $v' = 0$ level (see Appendix, Sec. 2, Fig. 19). As indicated in Table I (p. 506) this atmospheric band system of O_2 has been identified by Meinel in the infra-red night air-glow spectrum.

Using pure N_2, Kaplan obtained afterglow spectra the earlier parts of which resembled the nitrogen spectrum of the auroral afterglow and the later parts that of night air-glow. The essential discharge condition for producing these afterglows is that the relative intensity of the first negative bands of N_2^+ in the discharge must be very high [69]. The afterglows can then be obtained at all pressures ranging from 150 mm. to the lowest pressure at which discharge is possible.

The most important characteristic of the afterglow is that the relative and absolute intensities of its spectral components change with time in course of the decay of the afterglow which is about 10 seconds. At the beginning, the first negative bands are strong, stronger than the second positive bands. The Vegard-Kaplan bands and the forbidden atomic nitrogen line λ 3467 ($^2P \rightarrow {}^4S$) are also present, but are weak. (See Appendix, Sec. 2, Figs. 15 and 17.) The spectrum at the beginning is thus a very faithful reproduction of

the auroral spectrum so far as N_2-bands and N-lines are concerned. Thereafter, though the first negative bands steadily decrease in intensity, the first positive bands, the Vegard-Kaplan bands and the nitrogen lines gain in both relative (to the first negative bands) and absolute intensity. Towards the end of the glow the Vegard-Kaplan bands are found to have increased in intensity relative to the nitrogen lines, and the first negative and second positive bands are extremely weak. The spectrum at the end is thus a reproduction of the night air-glow for the nitrogen bands.

If oxygen is introduced in the discharge tube the green atomic oxygen line in the visible and the atomic nitrogen line (λ 3467) and the atomic oxygen line λ 2972 ($^1S \rightarrow {}^3P$) in the ultraviolet are increased greatly in intensity.

As already mentioned, the nature of the conditioning of walls of the discharge vessel (which prevents reactions on the walls and makes the afterglow possible) is still ill understood. (See Ref. [64]). However, the spectral emissions and the changes in their intensities with time (at least for the case of pure nitrogen) may be understood as follows after Mitra [70], if the very plausible assumptions are made that the primary source of energy in the afterglow is the excited N_2^+ (A') ions (which had been produced in the discharge) and that these excited ions have long life. The long life is evident from the persistence, for several seconds, of the first negative bands [$N_2^+(A') \rightarrow N_2^+$ (X')] after the discharge has been cut off. (In the alternative, one has to assume that the excited N_2^+ (A') ions are produced in the afterglow. But this is extremely unlikely.)

At the start, in the first part of the afterglow, there is naturally high concentration of excited N_2^+ (A') ions derived from the discharge. Hence the first negative bands are strong. With the progress of time, there will be accumulation of N_2^+ (X') ions in the ground state in the vibrational levels $v'' = 0, 1, 2$, etc. Now, according to Mitra [71], the ions in the vibrational levels ($v'' > 0$) and those in the lowest levels ($v'' = 0$) will recombine with electrons by two different processes. The first by a dissociative recombination process and the second by a three body collision process in which the third body is a N_2 molecule.

The first, the dissociative recombination process [73] is discussed in Chap. V, Sec. 3, and also in Chap. VIII, Sec. 6(c). In this process the N_2-molecule is dissociated producing two nitrogen atoms one in the ^2P state and the other in the ^2D state. (See Appendix, Sec. 2, Fig. 15.) The emission of the λ 3467 nitrogen line in the afterglow, by the transition $^2P \rightarrow {}^4S$ is thus explained. (The atoms in the ^2D state have the extraordinarily long life of about 8 hours. Hence they are de-excited by collision by a radiationless process.)

The second, the three-body collision process is discussed in the note above. In this process two excited nitrogen molecules, one in the B-state and another in the A-state are produced. (See Appendix, Sec. 2, Fig. 17.) The emissions of the Vegard-Kaplan bands ($A \rightarrow X$) and the first positive bands ($B \rightarrow A$) are thus explained. It should be noted in this connection,

that this process of neutralization of the N_2^+ (X') ions was proposed by Mitra [64], to explain the ordinary Lewis-Rayleigh afterglow (*vide supra*) in which *only* selected bands of the first positive system are present and no Vegard-Kaplan bands. The absence of the latter is explained by collisional de-excitation of the metastable N_2 (A) molecules on the walls of the afterglow tube. Their presence in the Kaplan afterglow is explained as due to the special conditioning of the glass walls of his discharge tubes in which de-excitation on the walls of both N_2^+ (A') ion and N_2 (A) molecule is prevented.

The above processes leave unexplained the presence of the second positive bands in the Kaplan afterglow. A number of suggestions may be made for this. For example, there may be three-body collision involving N (^2D) atoms which must have a high concentration on account of long life.

$$N(^2D)+N(^2D)+N_2 \rightarrow N_2(C)+N_2,$$

or, radiative recombination of N_2^+ (X') ions

$$N_2^+(X')+e \rightarrow N_2(C)+h\nu.$$

From the discussion given above the main features of the intensity change with time of the lines and bands are easily understood. The after-glow starts with a high intensity of the first negative bands as the initial concentration of $N_2^+(A')$ ions, derived from the discharge is high. With the progress of time the concentration of $N_2^+(X')$ in the ground state ($v''=0$, 1, 2, etc.) increases. Hence, the lines and bands emitted, in course of the two recombination processes described above, increase in intensity both relatively and absolutely. Towards the end of the glow, those of the excited molecules which have long life will be expected to have a relatively high concentration. These are N (^2P) atoms and N_2 (A) molecules. Hence, the emissions from these are relatively strong towards the end of the afterglow.

In view of the discussion above one may say that the active substance in the Kaplan glows is the excited $N_2^+(A')$ ion and that in the Lewis-Rayleigh glow is the $N_2^+(X')$ ion in the lowest vibration level ($v''=0$). Further, in the Lewis-Rayleigh afterglow tube the walls are so conditioned that only recombination of ions and electrons on the walls is prevented. In the Kaplan afterglows in addition, the de-excitations of $N_2^+(A')$ ion and excited $N_2(A)$ molecule on the wall are prevented.

5. NIGHT AIR-GLOW—PHOTOMETRIC AND POLARIZATION STUDIES

(a) Introduction

Measurement of the intensity of the night air-glow (for selected spectral regions or for the overall radiation) is important for many reasons. Thus, the measurement of the absolute intensity of a line radiation enables one to make an estimate of the corresponding number of emitted quanta, a knowledge of which is important in the theory of the excitation of the

line. Again, simultaneous measurements of the intensity at different zenith distances provide data for estimating the height to the glow-emitting layer.

Photometric measurements show that the brightness of the air-glow in the different parts of the sky is not uniform. There are limited regions or patches where the brightness is greater than the average. (The regions near the horizon are generally brighter than those near the zenith. But this is simply due to there being more luminous matter in the line of sight directed towards the horizon than towards the zenith; cf. Sec. 6(*a*)).

Further, the brightness not only varies from day to day—some nights being of exceptional brilliance—but also, it varies progressively in course of a single night. Superposed on these fluctuations are slow variations, semi-annual and annual. It has also been found that the average brightness of the air-glow waxes and wanes with the 11-year solar cycle. We may classify the temporal variations of the brightness of the air-glow as follows:—

(1) Slow variation related to solar cycle.
(2) Annual and semi-annual variations.
(3) Diurnal variation.
(4) Irregular fluctuations.

These variations are similar to those exhibited by the electron concentration of the ionospheric regions [Chap. VI, Sec. 11] and there is strong evidence that the two upper atmospheric phenomena—ionization and night air-glow may be physically related to each other [see Sec. 7].

In what follows we will first give brief descriptions of the methods of making the photometric measurements.

(*b*) Methods of measurements

Photometric measurements of the brightness of the air-glow—total brightness or brightness due to any particular region of the spectrum—may be made visually, photographically, or with photo-electric photometers.

For visual measurements a simple photometer (Fig. 18) as devised by the French school of workers may be employed [4, 98]. The observer looks at the lens *L* with his eye at the aperture *O*, the distance between the lens and the aperture being equal to the focal length of the lens. One-half of the lens is illuminated by the light from the sky and the other half by diffused light coming from the screen *D*. The screen *D* is illuminated by light from the lamp *S*. A microscope eye-piece L_2 of short focal length ($f = 25$ mm.) forms a small image of *S* in the plane *C* where an absorbing wedge is placed. It is clear that the observer may make the two halves of the lens *L* uniformly illuminated by adjusting the intensity of the light falling on *D* with the help of this wedge. In order to produce the same tint in both the halves of the lens, the light from the lamp *S* is made to pass through a filter of a dilute solution of ammoniacal copper sulphate. The precision of the measurement with this instrument is quite good when the angle subtended by the illuminated surface is large enough (0·1 to 0·2

radian) and the brightness is not below 10^{-8} candle per cm.², which is about a hundred times more than the minimum of perceptible brightness.

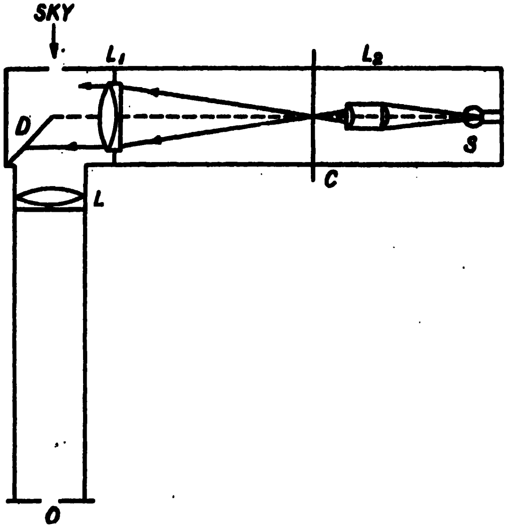


FIG. 18. Visual photometer as devised by Dufay.

Lord Rayleigh and his collaborators have used successfully filters for visual photometry of the three spectral regions—red, green and blue of the night air-glow [74].

For the purpose of comparison the light transmitted through the filter was equated to an artificial self-luminous source (potassium-uranyl sulphate) which is fluorescent due to the radioactivity of uranium. Neutral tinted glasses were used to match the brightnesses. As the glasses had densities ($\log_{10}$ opacity) 0·1, 0·2, 0·3, etc., there was an increase of one unit of brightness from one glass to the next and the intensities could be measured in terms of positive or negative numbers. Brightness zero means that no neutral glass is necessary to secure a match, while $+1$ or -1 (for example) means that glass number one has to be used over the sky or over the artificial source to secure a match.

The equipment for photographic measurement in its simplest form may consist of a photographic plate exposed behind an aperture, say, 1 cm. diameter made in a black paper and separated from it by a wooden board of thickness 1·5 to 2 cm. The plate would thus receive light from a cone of about 14° of the sky. The brightness of the sky observed may easily be obtained from the blackening of the photographic plate when the same has been previously calibrated. With such an apparatus, an exposure of 10 mins. may suffice to produce sufficient blackening—provided suitable photographic plate is chosen. For the photometry of isolated spectral regions suitable filters may be used.

An improvement of the above apparatus is to employ an object glass to gather more light as shown in Fig. 19. L_1 is an objective of focal length

50–60 cm. A circular opening, the size of which may be varied, is placed in its focal plane C. This opening limits the region of the sky observed. The

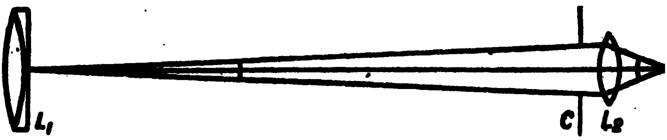


FIG. 19. Photographic method for measuring the brightness of the night sky.

lens L_2 behind the screen has a large aperture (about $f/1{\cdot}5$) and it throws on the photographic plate P, an image of the exit pupil of L_1. It is clear that C being in the focal plane of L_1, the image on the photographic plate will have uniform illumination. If necessary, neutral tinted screens of different densities may be placed before the objective L_1 to regulate the illumination.

The photoelectric photometer in which, as the name implies, a photoelectric cell is used for recording the intensity is the most sensitive and accurate instrument. It is also extremely useful in studying short period variations of the air-glow brightness. The K-H photo-cell is sensitive in the blue-violet region (Vegard-Kaplan bands), and the Cs-O cell in the infra-red region. K-O cells with filter have also been successfully used for measurement of green-blue region [75]. Elvey and Roach have employed photoelectric photometry for making a detailed study of the brightness of the night air-glow [13].

Photoelectric photometers have also been specially designed and developed for quick measurements of intensity at different altitudes and azimuths (as necessary for measurement of the height to the glowing layers). A description of such an apparatus, as used by Roach and Barbier [76], is given below. For fuller detail ref. [77] should be consulted.

The sensitive element of the photometer is an RCA–1 P 21 multiplier tube. It receives light from a two-element lens of four-inch aperture and four-inch focal length. The patch of the sky to be observed (2° by 10°) is limited by a diaphragm in the focal plane. A chopper in the conveying light beam produces a 30 cycle/s signal. The multiplier tube increases the initial photoelectric current by a factor of about one million. The resulting current is further amplified by a two-stage A.C. amplifier, the maximum current amplification of which is about 10^5, the sensitivity being controlled by the choice of a number of grid leak resistors in the second stage. The recording meter has a range of 0 to one milliampere. The telescope is on an alt-azimuth mounting so that it can be trained to any part of the sky. For any fixed azimuth it can sweep from horizon to the opposite horizon and repeat the performance indefinitely. The azimuth-setting can be changed by 22½° at the end of each sweep and thus the whole sky can be covered in a series of eight horizon to opposite horizon sweeps. A single sweep requires 4 minutes and a complete survey in a series of 8 sweeps—32 minutes. The spectral region (in the visible) to be studied is selected

by means of an interference type filter (manufactured by Baird Associates) which passes a band of about 150 A at half intensity. Two filters (2 inches by 2 inches) are used; one of these is a control filter admitting a spectral region (λ 5210) in which there is no prominent night radiation and the other including the radiation to be studied (λ 5577 oxygen or λ 5893 sodium). A mechanical shutter placed in front of the lens admits radiation alternately through one filter and then through the other filter at 4 seconds interval.

A similar photoelectric photometer in which the sensitive element is a gas-filled photoelectric cell (type GM 71–TA, manufactured by the GM Laboratories) has been used by Roach, Pettit and Williams for studying the infra-red region 6400 A to 11000 A [78]. A Corning filter, Number 241 was used to exclude all radiation shorter than 6400 A. The sensitivity of the arrangement was a maximum between 7000 A and 8000 A (Height measurement results as were obtained with these photo-electric photometers, will be described in Sec. 6).

Measurements of night air-glow intensity with photoelectric photometer have also been carried out in Japan [89].

(c) Annual and solar cycle variations—latitude effect

These variations are best illustrated by the results of observation and analysis as made by Rayleigh and Spencer Jones [74] the pioneer workers in this field, during the years 1923–1934 at the three stations Terling, England (43° from north magnetic pole, 1923–1934), Canberra, Australia (38° from south magnetic pole, 1925–1933) and at Cape Town, South Africa (69° from south magnetic pole, 1925–1933). The measurements were made visually in three spectral regions, blue, green and red by means of filters as described in the preceding section. The most prominent bands in the blue region were the Vegard-Kaplan bands; the green region included the green oxygen line and the red region comprised the first positive bands of N_2, the red oxygen lines and the band at λ 6580 now known to be due to the OH radical.

The observational results were subjected to harmonic analysis, from which two periodic terms—one annual and another semi-annual were determined. These are depicted in Figs. 20 and 21. The variations of the mean yearly values of the intensities together with the variation of the mean sunspot areas for the corresponding period are shown in Fig. 22. The following interesting features of the curves may be noted:

(i) The amplitude of variation of the green radiation is the largest. Next in order are the red radiation (in which is included the forbidden red radiation of atomic oxygen) and the blue radiation (in which the strongest are the Vegard-Kaplan bands). These lesser amplitudes may perhaps be attributed to the longer life of the metastable ^1D state of the O-atom, and of the A-state of the N_2 molecule respectively from which these radiations orginate.

(ii) The amplitude of variation for all the three regions is smallest for Cape Town which is farthest from the magnetic poles. However, if the

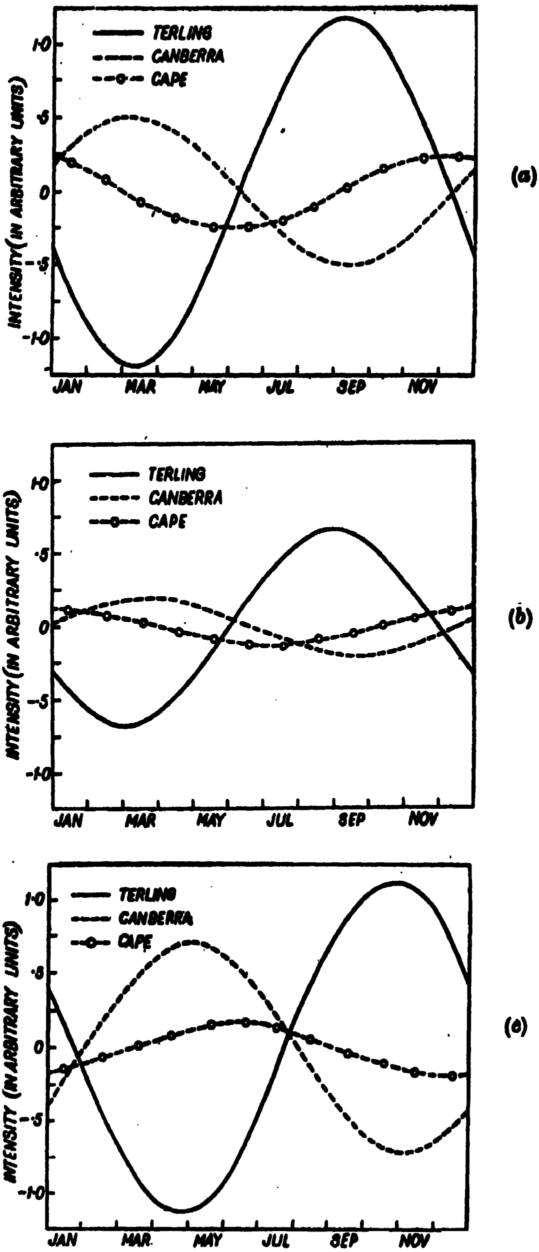


FIG. 20. Annual periodic terms obtained from harmonic analysis of the variations in intensity of the night air-glow spectrum at the three stations, Terling (England, 43° from the north magnetic pole), Canberra (Australia, 38° from the south magnetic pole) and Cape Town (South Africa, 69° from the south magnetic pole). (*a*) is for the green oxygen line, (*b*) is for the blue region (Vegard-Kaplan bands) and (*c*) is for the red region (oxygen lines and N_2 first positive and OH-bands)

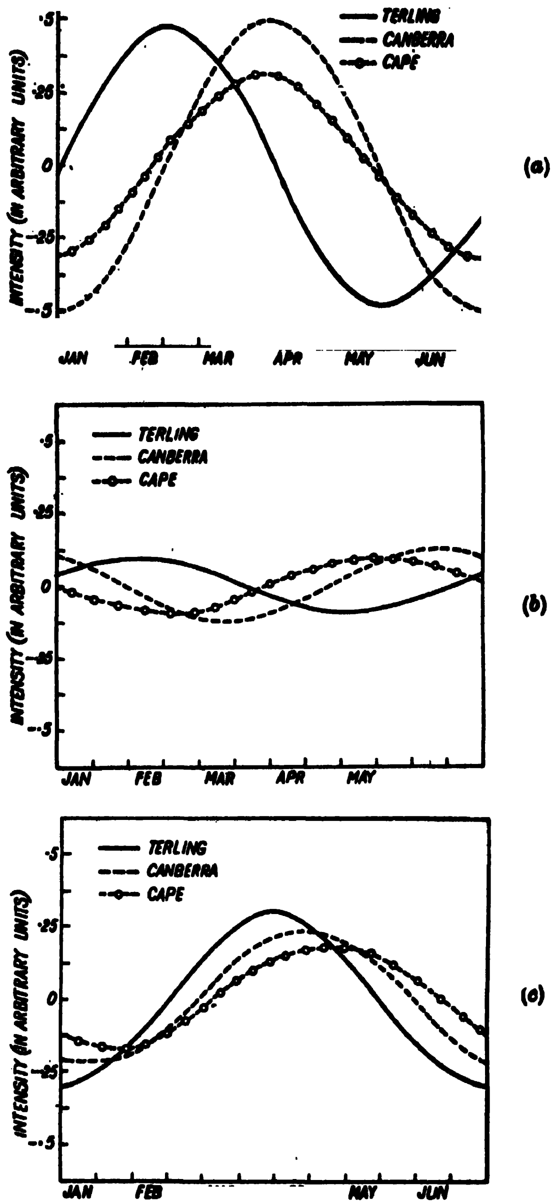


FIG. 21. Semi-annual periodic terms obtained from harmonic analysis of the variations in intensity of the night air-glow spectrum at the three stations, Terling (England, 43° from the north magnetic pole), Canberra (Australia, 38° from the south magnetic pole) and Cape Town (South Africa, 69° from the south magnetic pole). (*a*) is for the green oxygen line, (*b*) is for the blue region (Vegard-Kaplan bands), and (*c*) is for the red region (oxygen lines and N_2 first positive and OH-bands).

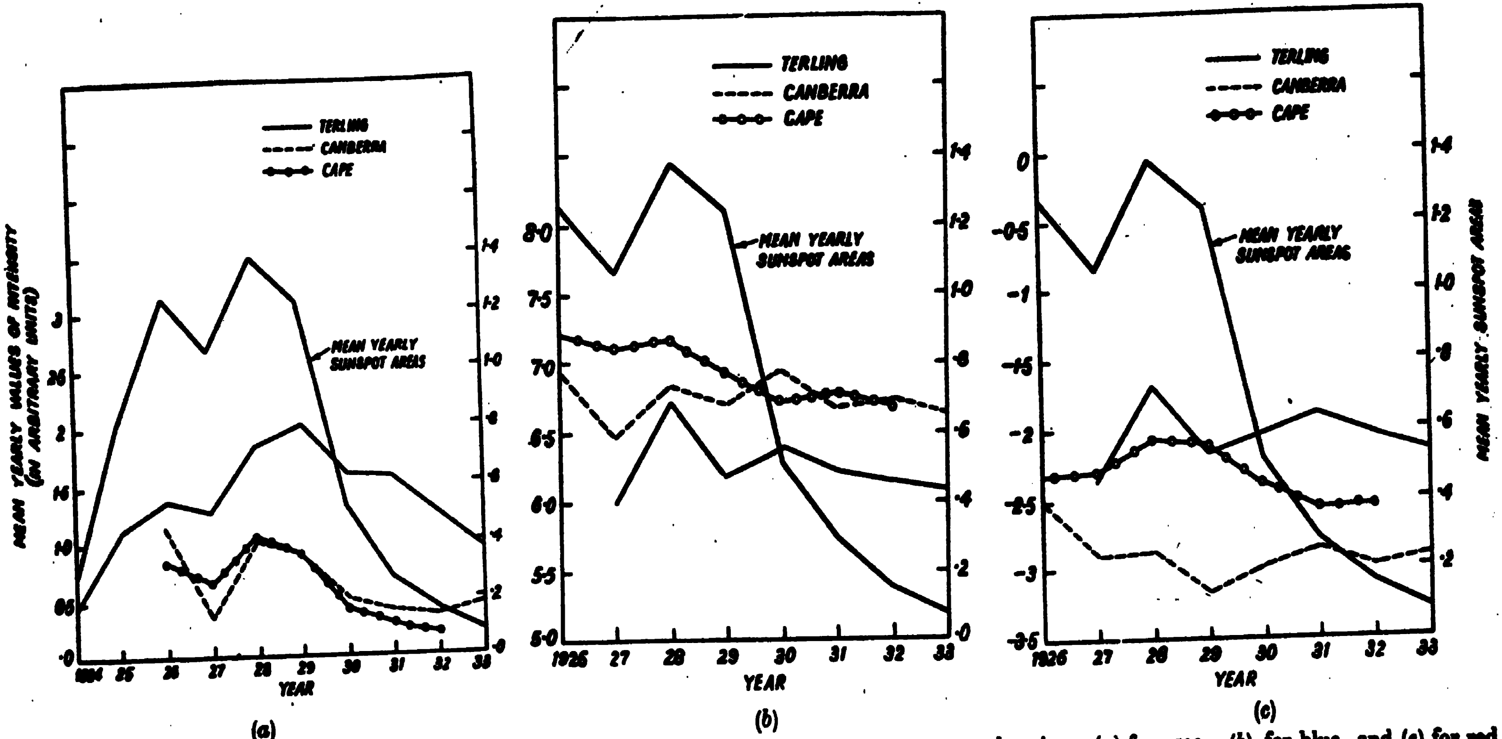


FIG. 22. Illustrating the variations of the mean yearly values of the intensities of the three spectral regions: (a) for green, (b) for blue, and (c) for red at the three stations mentioned in Figs. 21 and 22. The variations are compared with the mean yearly sunspot areas. (After Rayleigh and Jones.)

distance from the magnetic pole alone determined the strength of the variations, then Terling amplitude should have been lower than the Canberra amplitude. Actually the reverse is the case. This shows that besides the magnetic latitude the geographic location of the place is also concerned.

(iii) For a given station the variations of the three spectral regions are very approximately in phase. But, there are marked differences in phases for a given spectral region between the three stations. The Terling and the Canberra variations are nearly in opposite phase.

(iv) The variations in the amplitudes of the semi-annual term (Fig. 21) from station to station are much smaller than those of the annual term.

(v) The variation of the mean yearly value of the green radiation (Fig. 22(*a*)) shows the most marked correspondence with the variation of the mean yearly sunspot areas. However, the amplitude of the intensity variation is much less than that of the sunspot area variation. The variations for the red and the blue regions are again much less.

Results of observations by other workers also generally confirm those of Rayleigh and Jones, viz., that there is a semi-annual period superimposed on an annual period [79, 80], that at a given station there is parallelism between the variations of the intensities of the oxygen lines and the Vegard-Kaplan bands [81, 81(*a*)], that the intensity increases as one moves away from the equator [81(*b*), 81(*c*), 81(*d*)] and that there is a general tendency for the intensity to increase with the spottedness of the sun, the change in the green line being quite marked and that in the red lines and N_2 bands very little. It is to be added that the mean intensity of the sodium *D*-lines has also been found to undergo annual variation like the oxygen lines. According to Dufay's observations [81(*a*)] the intensity in summer is five times feebler than in winter. But, on individual occasions it can be 40 times as feeble.

The variation in intensity of the spectral region yellow to near-infra-red ($\lambda\lambda$ 5800–9000) has been specially studied by Grandmontagne [82]. He reports three maxima (instead of two of the other workers in the visible region), one in November, one (smaller) in March and one about June-July.

It is interesting to note that Dufay [81(*a*)] reports a 27-day recurrence tendency in the intensity variation of the green oxygen line.

(*d*) Diurnal variations

The nightly variation of intensity of the spectral components of the night air-glow has been studied in different parts of the world. The interpretation of the observed results is complicated by the frequent presence of irregular variations (*vide infra*) superposed on quasi-regular nightly variations. It is, therefore, necessary to consider observations on 'quiet' nights.

The quasi-regular variation may consist of a slow fall and rise with a minimum round midnight [55, 86, 87], or, a rise in the earlier part of the night and then a fall after attaining a maximum round the same hour [83, 84, 85]. According to Elvey and also Dufay [86, 81(*a*)] there are also quiet

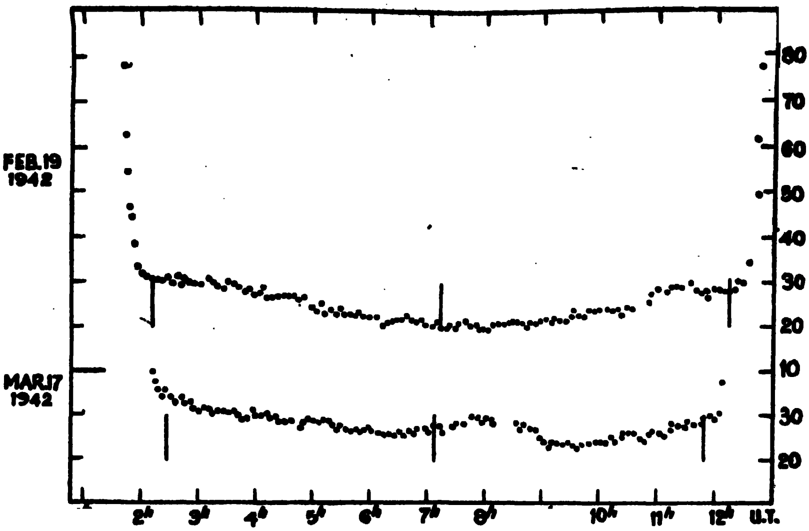


FIG. 23(a). Illustrating the type of variation in which the intensity of the night air-glow decreases as night progress, attains a minimum near about the local midnight and then rises again. The variations are interpreted to be due to the red OH-bands. (After Elvey.)

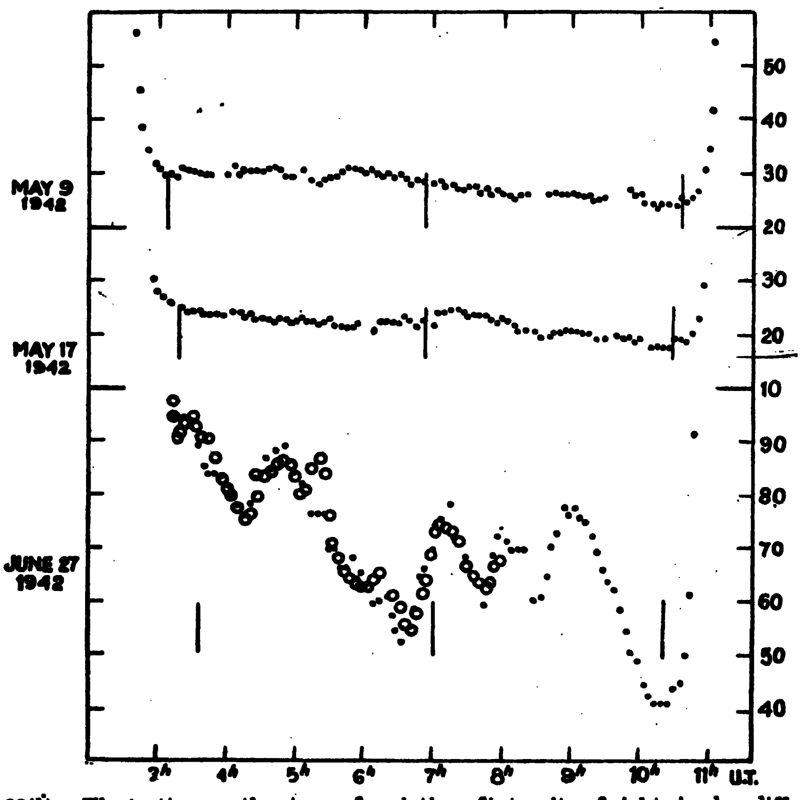


FIG. 23(b). Illustrating another type of variation of intensity of night air-glow different from that depicted in Fig. 23(a) (the two top curves). There is a steady decrease in intensity throughout the night. The lower curve is for a disturbed night and shows strong irregular fluctuations. (After Elvey.)

nights in which the intensity falls steadily with the progress of the night. In Fig. 23(*a*) and in the upper part of Fig. 23(*b*) the nightly variations of the types, midnight minimum and steady fall, are shown after Elvey [86]. (The sharp rises in the early morning and in late evening are twilight effects. See Sec. 7(*b*)). In Fig. 24 the type of variation with midnight maximum is shown after Roach and Pettit [95]. Such a variation has also been observed by Dufay. It appears that the two types of variation—midnight minimum and midnight maximum—refer to emissions from different heights. Thus the curves in Fig. 24 (midnight maximum) are for the green

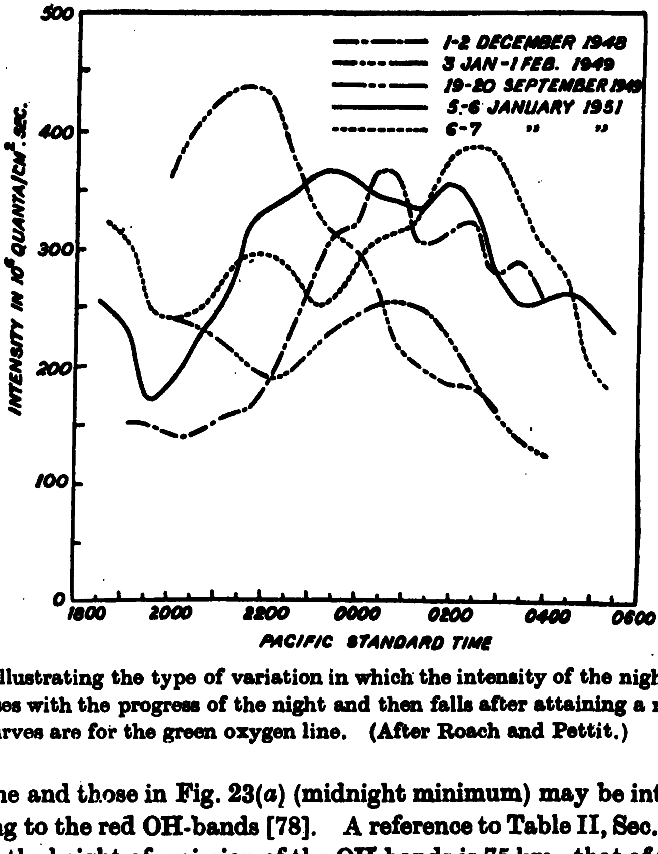


FIG. 24. Illustrating the type of variation in which the intensity of the night air-glow increases with the progress of the night and then falls after attaining a maximum. The curves are for the green oxygen line. (After Roach and Pettit.)

oxygen line and those in Fig. 23(*a*) (midnight minimum) may be interpreted as referring to the red OH-bands [78]. A reference to Table II, Sec. 6, shows that while the height of emission of the OH-bands is 75 km., that of the green oxygen line is 250 km. (Roach and Pettit, Karimov).

Superposed on the quasi-regular variations described above there are irregular variations as depicted in the lower part of Fig. 23(*b*) after Elvey. These we now proceed to discuss.

(*e*) Irregular fluctuations

It has been noticed by all observers that there are nights characterized by unusual brightness of the sky. These are not attributable to auroral displays. According to Karandikar [87] the variations of intensity on such

nights are also abnormal and the variations are analogous for the different spectral regions.

According to Roach and Pettit there is present in the upper atmosphere fixed excitation pattern on the dark side of the earth, through which a given

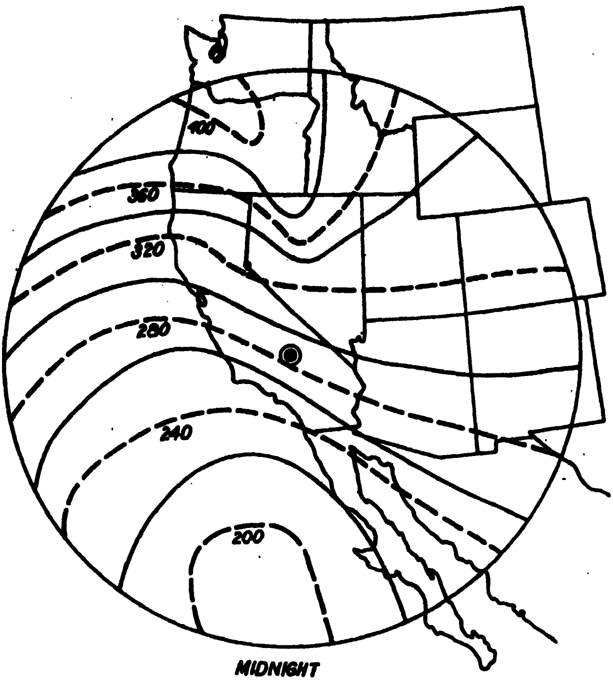


FIG. 25. Isophotal plots of the fixed excitation pattern for the green oxygen radiation as existed over Cactus Peak (centre of the circle), Mojave Desert, South California, at midnight on January 6/7, 1951. The map is based on a height of emission of 250 km. The figures multiplied by 10^6 give the intensity of emission in number of photons emitted per $cm.^2$ column per sec. (After Roach and Pettit.)

observer seems to move with the rotation of the earth. Fig. 25 shows the portion of the pattern (isophotal plots) for the green oxygen radiation as existed over the observing station, Cactus Peak, Mojave desert, South California on January 6/7, 1951, at midnight [95].

It is now established that irregular fluctuations of night air-glow intensity are closely related to magnetic activity. This is in accord with the observations of Rayleigh and Spencer Jones (*vide supra*), viz., that the mean intensity is greater in higher latitudes, because, the magnetic activity is known to increase as one moves towards the poles.

The interesting curves in Figs. 26 and 27 are due to Barber. He has compared seasonal and diurnal variations of the intensity of the night air-glow as observed at Hamilton, California with magnetic activity as recorded at Mt. Wilson situated at a distance of 300 miles. It will be noticed in Fig. 26 that the seasonal variations of the two follow each other closely. In Fig. 27 the diurnal variations of the two are compared. Here, again, there is close

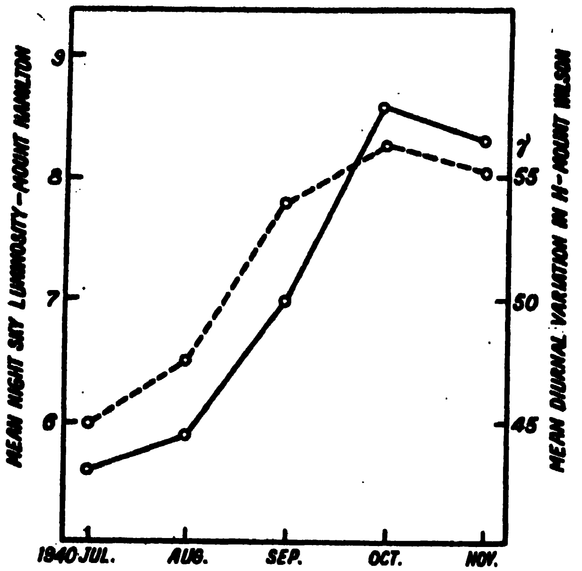


FIG. 26. Comparison of the seasonal variations in the intensity of the night air-glow (——) with the mean diurnal variation in H(- - -). The former is recorded at Mt. Hamilton and the latter at Mt. Wilson at a distance of 300 miles to the south-east. The night sky luminosity is given in arbitrary scale. (After Barber.)

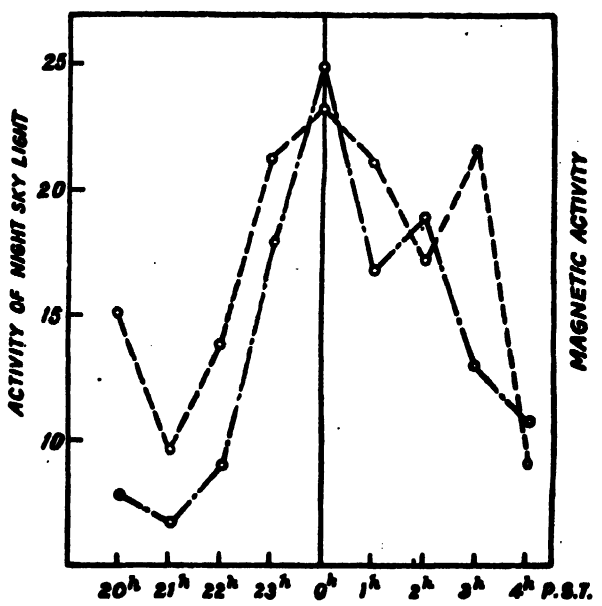


FIG. 27. Comparison of the diurnal variation in night sky activity with that of magnetic activity. The former is recorded at Mt. Hamilton and the latter at Mt. Wilson 300 miles south-east. Activity of the night sky light is the product of hourly mean and hourly range of luminosity. (After Barber.)

parallelism between the two curves. It is to be noted that in both cases the disturbed hours tend to group around midnight.

Some interesting observations on the gradual shift of the phase of the intensity variation in course of a night have been reported from Japan [89]. The observations were made on the OH band at λ 10450 and on the red oxygen lines at λ 6300 in three directions, viz., zenith and 35° and 20° altitude north. It was found that there was a gradual shift of the phase of the variation from north to south. Assuming that the height to the emitting layer was 300 km. (*vide infra*), the velocity of the shift was found to be about 220 m/s. These observations were in conformity with the observations on the movements in the ionospheric regions (E_s for example) from north to south made in the same country [90]. The author, however, is of opinion that the shift is to be explained rather as due to movement of some influence on the night sky than to movement of any material of the upper atmosphere.

(*f*) Absolute intensities of the spectral radiations

The pioneer measurements on absolute intensity are those of Lord Rayleigh [91] who obtained for the green and the red atomic oxygen radiations the values 2×10^8 and 1×10^7 transitions per second per sq. cm. column respectively. These results have generally been confirmed by later measurements [92].

Measurements of the sodium *D*-lines have yielded values between 2×10^7 and 8×10^7 transitions per sq. cm. column per second [57, 94, 123].

Measurements on OH band systems have been carried out with photoelectric photometer sensitive from 6400 A to 11600 A [78]. The average absolute intensity is found to correspond to $5{\cdot}8 \times 10^9$ transitions per sq. cm. per second. The total intensity integrated in the entire spectral region under consideration was estimated to correspond to 4×10^{10} transitions per sq. cm. column per second.

It is to be remembered, however, that the intensities of all the spectral components of night air-glow fluctuate within wide limits specially on disturbed nights. The above may be taken as representative values for quiet nights.

(*g*) Polarization measurements

The light of the night air-glow has been found to be partially polarized; the proportion of polarized light is, however, quite small—two to four per cent only. The plane of polarization is found to pass constantly through the sun. Very precise measurements of polarization have been made by Dufay [20] with particular reference to the polarization of the zodiacal light. The amount of polarization, which may be estimated from the depolarization factor varies slightly with the progress of night, i.e., with the position of the sun with respect to the region of the sky observed. [The ratio between the minimum and the maximum of light transmitted through the analyzing nicol when it is rotated round the line of sight is called the depolarization

factor ρ. The proportion of polarized light is given by $(1-\rho)/(1+\rho)$.] Fig. 5, upper curve shows after Dufay the variation of depolarization factor with the angular distance from the sun. The region of the sky observed was in the meridian towards the north. It will be noticed that there is a slight depression of the curve when the angular distance of the sun is about 65°. The same author has also measured the depolarization factor for the zodiacal light in the region of the ecliptic. The results are plotted in the lower curve of the same figure. It is at once noticed that the zodiacal light is more strongly polarized than the light of the night air-glow. Further, the general trend of its variation is similar to that of the night air-glow. It may be suggested for this reason that the polarized component of the night air-glow is due to an extension of the zodiacal light. [See also Sec. 2(*a*).]

Khvostikov [96, 97] made some observations during 1938-1939 on the polarization of the oxygen lines λ 5577, λ 6300 and the sodium line λ 5890, at Siméise, Crimea. He found the following results amongst others. The degree of polarization varies between 9 to 17 per cent. The variation is similar in the three cases. The plane of polarization is always in the region of the direction perpendicular to the plane passing through observer, the pole and the sun. The measurements were made in the region of the Pole Star.

6. HEIGHTS OF THE LUMINESCENT LAYERS

(*a*) Van Rhijn method

The method of measuring the height of a luminescent layer is based on the assumption that the thickness of the layer is small compared to the radius of the earth, and that its luminosity is uniform over large geographical distances [97(*a*)]. It is obvious that the intensity of light received from such a layer, from any direction, would be proportional to the amount of luminous matter traversed by the line of sight in that direction. Thus, if the layer is assumed to be horizontal and flat (neglecting the curvature of the earth), the zenithal intensity I_Z and the intensity I_θ in a direction inclined at an angle θ to the vertical will be related by

$$I_\theta = I_Z \sec \theta.$$

If the curvature of the layer is taken into account then,

$$I_\theta = I_Z \sec \alpha$$

where α is given by,

$$\sin \alpha = \frac{a}{a+h} \sin \theta,$$

a and h being respectively the radius of the earth and the height of the luminescent layer. Thus, the ratio of the intensity in the direction of the zenith to that in some direction, at zenithal distance θ, enables one to compute h, the height to the emitting layer. We note that the intensity

ratio I_θ/I_z increases from unity at zenith, to a finite value greater than unity near the horizon—the variation with zenith distance being a function of height. (A thin layer has been assumed; but it can be shown that within reasonable degree of accuracy the method gives the mean height of a thick layer [36].)

(b) Extinction corrections

Important corrections have, however, to be applied to the simple formula given above, because, the sky intensity as observed does not represent the true brightness of the sky in the region of observation. The light received suffers 'extinction' in its path and to it is added the light scattered from the lower atmosphere. As a result of the former the observed intensity, instead of increasing with increasing zenith distance (as according to the simple formula given above) actually decreases near the horizon

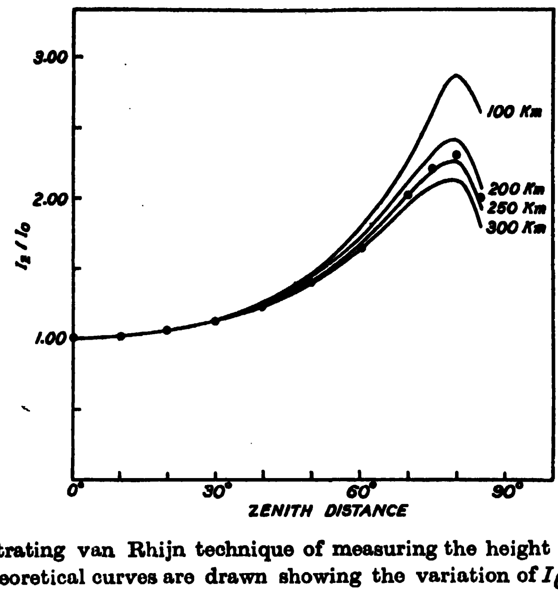


FIG. 28. Illustrating van Rhijn technique of measuring the height of a luminescent layer. Theoretical curves are drawn showing the variation of I_θ/I_z with zenith distance (after making corrections for extinction and scattering) for different assumed heights of the emitting layer. The height corresponding to the particular curve which fits best with the observed curve is the height of the emitting layer. The points lying on the 250 km. curve were obtained by Roach and Pettit [95] for the green oxygen radiation. Hence, the height of the corresponding emitting layer is 250 km. (*Note:* Read I_θ/I_z for I_z/I_0 in the ordinate scale.)

going through a maximum at a zenith distance of about 70° to 85° depending on the height of the emitting layer. As a result of the latter (contribution by scattered light) there is increase in the observed intensity with increasing zenith distance. This partly compensates the effect of the former, though the general shapes of the curves remain unaltered. (See Fig. 28.)

Barbier has worked out a method for eliminating the effects of absorption and scattered light. His original method is described in ref. [57]. A method developed later which is claimed to be simpler and quicker is as follows [99].

The apparent brightness I_θ^A of the sky in the direction of zenithal angle θ may be written as

$$I_\theta^A = KI_\theta + I'_\theta.$$

Here I_θ is the true brightness, KI_θ that reduced by atmospheric extinction and I'_θ is the contribution to I_θ^A by light scattered by the lower atmosphere into the line of sight.

Remembering that $I_\theta = I_s \sec \alpha$, the ratio of the apparent brightnesses for two directions θ_1 and θ_2 may be written as,

$$\frac{I_{\theta_1}^A}{I_{\theta_2}^A} = \frac{K_{\theta_1} I_s \sec \alpha_1 + I'_{\theta_1}}{K_{\theta_2} I_s \sec \alpha_2 + I'_{\theta_2}} = \frac{T_1 \sec \alpha_1 + I'_{\theta_1}}{T_2 \sec \alpha_2 + I'_{\theta_2}}.$$

(In the last expression T_1 and T_2 have been put equal to $K_{\theta_1} I_s$ and $K_{\theta_2} I_s$ respectively).

Now K_{θ_1} and K_{θ_2} (or T_1 and T_2) are obviously functions of the direction of observation (θ) and the height to the emitting layer (h). According to Barbier the scattered light term (I'_θ) can also be expressed as a function of θ and h involving the quantity T as, $I_\theta = (1-T)J(\theta, h)$ where $J(\theta, h)$ is of the form $F_a + F_b \cos^2 \theta$. The quantities F_a and F_b are independent of θ but vary with height and have been tabulated by Barbier. Thus, the ratio of the apparent brightnesses for two zenithal distances θ_1 and θ_2 (for any given radiation) may be written as

$$\frac{I_{\theta_1}^A}{I_{\theta_2}^A} = \frac{T_1 \sec \alpha_1 + (1-T_1)J(h, \theta_1)}{T_2 \sec \alpha_2 + (1-T_2)J(h, \theta_2)}.$$

The expression on the right is evaluated for given values of θ_1 and θ_2 (T_1 and T_2 being known) with different values of h and the values are plotted with $I_{\theta_1}^A / I_{\theta_2}^A$ as abscissa. The value of h in the plot which corresponds to the value of $I_{\theta_1}^A / I_{\theta_2}^A$ as observed gives the most probable height to the luminescent layer.

The above method of corrections for absorption and for scattered light has generally been adopted by the French school of workers.

Formula for extinction has also been proposed by Piotrowski [100]. Also, the extinction formula as proposed by Chandrasekhar [101] can be adapted for night air-glow intensity study. However, it has been shown by Barbier [102] that the results obtained with the help of these formulas do not differ greatly from those obtained with his formula, particularly, if higher approximations are taken.

In the U.S.A., Elvey and Farnsworth, who have carried out extensive measurements on night-air-glow intensity, have assumed arbitrarily for purposes of calculation, the effective value of the extinction coefficient to be one-half of the stellar extinction coefficient for the wavelength under consideration. According to these authors the many uncertainties involved do not warrant a more exact procedure.

Considering the fact that the luminous region is an extended source and the radiation is diffuse rather than collimated, Hulburt prefers for calculation of atmospheric absorption a formula due to Schuster [103] which is specially applicable to such cases rather than any exponential formula. Thus if β is atmospheric attenuation for collimated light, the attenuation of diffuse radiation by a scattering medium is given by $I_z^A = I_z/(1+\beta)$ and $I_\theta^A = I_\theta(1+\beta\gamma \sec\theta)$ where γ takes account of the curvature of the earth.

It may be noted that some observers also take into account the effect of absorption by the ozone layer (Dufay and Tcheng-Mao-Lin).

Estimation of the height to the emitting layers made even after corrections for extinction and scattering suffer, however, from a very grave handicap. The basic assumption in the van Rhijn type of formula is that the emitting layer is of uniform brightness. This assumption is not at all justifiable. Instead of a uniform brightness there are bright patches which, in all probability, move about. Under such condition the application of the van Rhijn type of formula becomes meaningless. For this reason some workers attempt to eliminate the uncertainty by taking a sufficiently large number of observations so that the effect of non-uniformity may be averaged out. Others have first determined the azimuth in which the luminosity appears to be most uniform and utilize readings only in this azimuth when making calculation. But even in such cases there is quite a large scatter in the calculated values.

For actual estimation of the height with the observed data the following procedure is often adopted. A set of theoretical curves, with I_θ/I_z plotted against zenith distance (after making corrections for extinction and scattering) for different assumed heights, is first drawn (Fig. 28). From these, the curve which agrees best with the observed curve is picked out and the height for which this curve has been drawn is assumed to give the height of the emitting layer. For the cases, in which I_θ/I_z is measured for only one fixed value of θ (round 80°), the average value, or the value which occurs most frequently is taken and the height calculated therefrom.

(c) Results of measurements

Results of some of the height measurements as made by the van Rhijn technique are given in Table II. (To avoid confusion the Table does not include the height measurements of sodium layer as made from observations of the twilight flash. This is discussed in Sec. 9.)

TABLE II.

Heights of the glow-emitting layers

Observer	Radiation	Height (Kilometre)	Remarks
Cabannes, Dufay and Gauzit (1938) [93]	Yellow (Sodium)	130	Measurements made in France.
Elvey and Farnsworth (1942) [36] ..	Red and green (Oxygen)	200 to 500	Measurements made with prism spectrograph at a number of stations in North and South America. Correction made for extinction; I_θ is for 5° above horizon.
Dufay and Tcheng-Mao Lin (1941–44) [105] ..	Green (Oxygen)	100	Measurements made with prism spectrograph at Haute Provence, France. Correction made for absorption by ozone in addition to that for extinction. Large scatter in the results.
	Red (Oxygen)	180	
	Yellow (Sodium)	80	
Vassy and Vassy (1943–46) [106] ..	Red (Oxygen)	750 100	Measurements made with photo-electric photometer at Pic-du-Midi, France. For the red there are two layers; the intensity of the lower is about one-tenth of the upper. For green there are also two layers; the intensity of the lower is about one-half of the former. Extinction corrections made.
	Green (Oxygen)	100 75 to 100	
Vassy, Abadie and Vassy (1949) [104]	Green (Oxygen) Red (Oxygen)	1,000 to 400	The height drops from 1,000 km. to 400 km. with the progress of the night.
Karimov (1947) [106(a)] ..	Green (Oxygen)	260	
Hulbert (1947) [107] ..	Total radiation	100	Measurements made with Macbeth illuminator and low brightness photometer at Brazil (17° 12·8′ S., 43° 41′ W.)
Roach and Barbier (1948) [94] ..	Green (Oxygen) Yellow (Sodium)	110 310	Measurements made with photo-electric photometer at Cactus Peak, Mojave Desert, U.S.A.
Roach, Pettit and Williams (1950) [129] ..	Infra-red; strongest radiations 10000 A to 11000 A (OH)	70	Same as above.
Roach and Pettit (1951) [95] ..	Green (Oxygen) Red (Oxygen) Yellow (Sodium)	250	Same as above.

7. EXCITATION PROCESSES—NIGHT AIR-GLOW AND TWILIGHT FLASH

(a) Night air-glow

The proposed theories for excitation of the observed lines and band systems in the night air-glow spectrum must not only provide for the necessary amount of energy at the required potential but also explain the observed variations and heights of the emissions. None of the theories proposed so far rigorously satisfy all these conditions. But, before discussing these theories it would be instructive to enquire into the ultimate source of energy which the air-glow represents.

(i) *Source of energy.*—It has already been mentioned that the intensity variation of the night air-glow consists of an irregularly fluctuating part and of a quasi-steady part. The former, as has been pointed out, shows strong correlation with magnetic activity and, as such, its excitation may be attributed to the same source as causes the aurora and magnetic disturbances, namely, charged particles shot off from the sun [88]. The charged particles move in curved paths under the influence of the magnetic field of the earth and, by bombarding the upper atmosphere in its dark side, produce patches of irregular illumination. As has been discussed in Chap. VIII the charged particles ordinarily concentrate in the so-called auroral belts situated 23° from the magnetic poles; but, during magnetic storms, they arrive in lower latitudes. The fact that the strength and the frequency of the irregular variations are related to the magnetic latitude of the place also confirms the view that their origin is the same as that of the aurora and magnetic disturbances.

As regards the quasi-steady background, the ultimate source of its energy must be the energy of the ultraviolet rays of the sun spent in ionizing and/or dissociating the upper atmospheric gases during daytime. After the withdrawal of the solar rays, the dissociated and the ionized products react amongst themselves reproducing the original bodies and in the process release the stored energy as radiation. An interesting question arises in this connection. How the stored energy is released slowly and is not dissipated quickly after the withdrawal of the exciting solar radiation? One recalls in this connection that the glow in a discharge tube, unlike the glow of the night sky, disappears within a fraction of a second as the source of excitation is withdrawn. The reason of this difference is not far to seek. For the case of the discharge tube recombination of the dissociated and the ionized bodies proceeds partly in the *volume* of the gas and partly on the *walls* of the tube. The rate of the latter is, however, very much faster than that of the former, so that practically all recombinations take place on the wall. Since in the upper atmosphere there is no 'wall-effect' the recombination proceeds only in the volume and is slow.

It will be interesting to inquire more closely into the part played by the wall of the discharge tube in promoting recombination, as this will give us some insight into the nature of the reactions which are likely to proceed in the upper atmosphere. Recombination of an electron and an ion is a complicated phenomenon because every approach

of an electron to an ion does not result in the capture of the former by the latter. The electron at a point near the ion possesses additional kinetic energy due to its free fall by attraction of the ion to the point under consideration. The electron therefore will be expected, on classical theory, to move past the ion in an hyperbolic orbit when it comes within close range of the attraction of a positive ion. In order therefore that the electron may settle down to a closed orbit, it is necessary for it to lose a part of its energy. This it can do either by radiation or by collision with a third body during the interval of its close approach.

In regard to the former—the so-called radiative recombination process—theoretical considerations show that though it has a finite probability, its coefficient is very small being of the order 10^{-12} cm.^{3}/sec. [109]. The radiation which occurs lies beyond the region of regular line spectrum and observations show that the intensity of such radiation is very feeble. Regarding the absorption of energy by a third body—the so-called three-body collision process—it is found that its probability is large only for comparatively high gas pressures. For such cases, the coefficient of recombination may be deduced from the formula of J. J. Thomson [110]. Approximate estimate (since J. J. Thomson's formula is applicable only to high pressures) shows that for pressures as encountered in the discharge tube (of the order of few tenths of mm. of Hg) the coefficient of recombination by three-body collision is extremely small. (See Chap. V, Sec. 2.)

An effective three-body collision may, however, occur in which the three bodies need not collide simultaneously, and for which the probability of recombination is much higher. For instance, an electron may first collide with a neutral particle (M) and attach itself to the same to form a negative ion (M^-). The negative ion so formed then collides with another neutral particle (N) or with a positive ion. In the first case the following reaction may occur

$$M^- + N \rightarrow MN + e \text{ (high energy).}$$

The two particles combine to form the molecule MN and the energy released is carried away by the electron as kinetic energy. The attached electron thus acts as the third body. Part of the energy may also be taken up by MN as energy of excitation, in which case the energy carried away by the electron is correspondingly less.

If the negative ion collides with a positive ion (X^+) then the following reaction may occur,

$$M^- + X^+ \rightarrow M \text{ (excited)} + X \text{ (excited).}$$

This is the process of so-called electron transfer in which both the reacting particles are neutralized and the energy of neutralization is taken up by the reaction products partly as excitation energy and partly as kinetic energy. The probability of the process is high if the whole of the energy is taken up as energy of excitation. (See, however, *infra*; next subsection.)

An alternative probability is the reaction

$$M^- + X^+ \rightarrow MX \text{ (excited),}$$

in which the two ions combine to form a molecule and the released energy is taken up as energy of excitation of the molecular product (MX). Here again the probability of the process is high only when the molecule (MX) possesses an excited state the energy of which is nearly the same as that liberated on recombination.

Simple calculations, taking into account the number of collisions which the reacting particles make per second show, however, that if the ions and the electrons were to disappear by any or all the above reactions occurring only in the *volume* of the gas in the discharge tube, then the rate of the recombination would be so slow that it would take seconds or even minutes for the ion density to fall to a small fraction of its original value. The glow in such case continues for an appreciable time instead of disappearing

almost simultaneously with the withdrawal of the exciting voltage. (*Vide* the note in Sec. 4(*d*).)

It is therefore evident that there must be some process other than the reactions proceeding in the *volume* by which the electrons and the ions disappear so quickly. This other process is reaction on the *wall* of the discharge tube in which the wall acts as the third body. The electrons striking the wall remain there as a surface charge; the ions arrive afterwards and combine with the electrons giving up the released energy to the glass wall. Since a solid has a very complicated structure and can take up the released energy in one of its many excited states the probability of the process is very high; practically every encounter of an ion with an electron on the wall is successful in producing the recombination. Further, since the number of electrons striking the wall is enormously larger than the number colliding with the charged or uncharged particles in the volume of the gas, the rate of recombination on the wall is very high indeed. The space density of electrons and ions in the discharge tube therefore falls to a negligibly low value, within a small fraction of a second, after the exciting voltage has been withdrawn, and the glow disappears.

(ii) *Atomic oxygen lines and* N_2 *bands.*—A simple process of ionic recombination between negative ions of oxygen atoms and positive N_2 ions by electron transfer, as indicated in sub-section (i) above, has been proposed by Mitra [111] and also independently by Nicolet [112] to explain the excitations of the O-lines and the N_2-bands:

$$N_2^+(X')+O^- \rightarrow N_2(B)+O(^1S) \qquad \ldots \qquad \ldots \qquad (I)$$

If the electron affinity of O-atom be taken as that given by Lozier [113], viz. 2·2 eV, then, since the ionization potential of N_2 for the $N_2^+(X')$ state is 15·58 eV, the energy released on neutralization is (15·58 − 2·2) eV = 13·38 eV. This energy is sufficient to raise the neutral N_2 molecule to the $v' = 9$ level of the B state (9·1 eV) and the O-atom to the ^1S state (4·2 eV). There is thus almost exact energy balance. The oxygen atom excited to the ^1S state drops successively to the ^1D and the ^3P states emitting the green and red lines (see Fig. 17). And, the N_2 molecule excited to high vibrational levels of the B-state drops successively to A-state and thence to the ground X-state emitting the first positive and the Vegard-Kaplan bands. (For energy level diagrams of O and N_2 see Figs. 16 and 17, Appendix, Sec. 2.)

A satisfactory feature of the hypothesis is that the one and the same reaction explains the emissions of both the O-lines and the N_2-band systems. In all other theories proposed two sets of reactions are suggested, one for the O-lines and another for the N_2-band systems.

It has been shown by Ghosh [116] (making plausible assumptions regarding the number densities of the particles involved) that according to process (I) the number of transitions per second in one sq. cm. column is $4{\cdot}9\times10^8$. This is in good agreement with the value as obtained by Rayleigh from measurement of the absolute intensity of λ 5577 radiation [115].

Mitra has suggested that the emitting region in the night sky is closely related to, if not actually identifiable with the F-layer of the ionosphere. In support of this, observational evidence is adduced that on disturbed nights, on which both $f_{\max.}\,F$ and the intensity of the night sky vary abnormally, the two variations follow approximately the same trend [117]

as shown in Fig. 30. It is to be noted, however, that too close a correlation for short period variations cannot be expected. This is because the ionospheric measurement gives only variations in the maximum number density of electrons. It does not say anything about the variation of the total number of electrons on which the intensity of the night air-glow depends. For example, the number density may increase due to contraction by cooling of the region as a whole. But the total number of electrons in a sq. cm. column in such case will remain more or less constant.

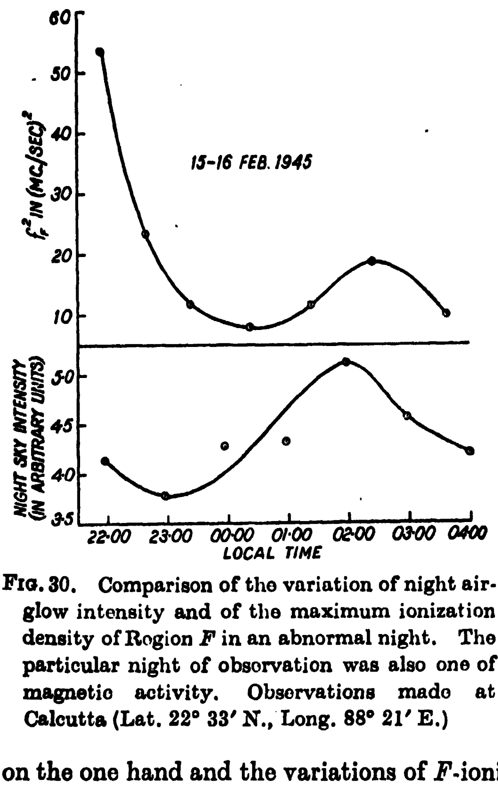


FIG. 30. Comparison of the variation of night air-glow intensity and of the maximum ionization density of Region F in an abnormal night. The particular night of observation was also one of magnetic activity. Observations made at Calcutta (Lat. 22° 33′ N., Long. 88° 21′ E.)

It may be mentioned in this connection that Japanese workers report that they have obtained correlation, though not very definite, between variations in the intensity of the OH bands (λ 10450) and the red and green oxygen lines on the one hand and the variations of F-ionization on the other [118]. For the OH band (*vide infra*) there was positive correlation and for the oxygen lines, negative correlation. It is not, however, stated if particularly disturbed nights of ionospheric storm were selected for observation as in the case of Calcutta observations.

It should be mentioned that criticisms of the theory of ionic recombination of Mitra as sketched above have been made on the two following grounds [119, 120].

A quantal study of mutual neutralization by ionic recombination [121] has shown that the most rapid of such processes are not those which have exact energy balance (as emphasized in Mitra's theory) but are those which are exothermic by a few volts. This conclusion has, however, been questioned by Bennett and also by Sayers [122].

The other objection is that the total number of N_2^+ ions in the upper atmosphere is too small to enable the reaction to proceed at the required rate [120].

As an alternative to the process proposed by Mitra one may consider the original suggestion by Chapman [124] in which the source of energy of the night air-glow is to be sought rather in the energy of dissociation of the

atmospheric constituents (O_2 and probably N_2), than in the energy of ionization as in Mitra's process.

For the excitation of the O-lines a three-body collision as follows may be possible

$$O+O+O \rightarrow O_2+O^*.$$

A similar suggestion cannot be made for the excitation of the N_2 bands because the probability of the analogous process is extremely low. The reason for this is that while in the O_2 molecule no electronic transition is involved, in the N_2 molecule it is not so [125]. The two nitrogen atoms in the normal ^4S state cannot approach each other along any potential energy curve from which there is allowed transition to a bound state (See Fig. 18, Appendix, Sec. 2). To obviate this difficulty Ta-You Wu [114] suggests that one of the colliding atoms might be in the long-life metastable state ^2D. However, this needs an impossibly large concentration of such atoms. It is, however, to be noted that if Gaydon's value of dissociation energy (9·76 eV) is adopted [139], then there is at least energy balance for the following three body collision process

$$N+N+N_2 \rightarrow N_2(X)+N_2(B),$$

in which one of the N_2 molecules on the right is raised to a high vibrational level of the B state ($v' = 8$, 9 or 10) and the other to some low vibrational level of the ground state N_2 (X). From the former the first positive bands ($B \rightarrow A$) and then the Vegard-Kaplan bands ($A \rightarrow X$) are emitted.

(iii) *Herzberg and atmospheric bands of* O_2.—A three body recombination of O-atoms as follows:

$$O+O+M \rightarrow O_2^*+M,$$

has been suggested for the excitation of these bands. For three body collision processes the coefficient is of the order 10^{-32} cm.^{6}/sec. If in the process electronic transition is involved then the rate is much smaller. However, for the above relation the coefficient needs be small.

(iv) *D-lines of sodium.*—A cyclic process is generally proposed in which sodium atoms form sodium compounds (NaH or/and NaO) and the compound is then dissociated by collision with O-atoms yielding excited sodium atom [124].

Regarding the formation of the sodium compound various reactions may be imagined, e.g.,

$$Na+O_2+M \rightarrow NaO_2+M$$

and then

$$NaO_2+O \rightarrow NaO+O_2.$$

also/or

$$NaO_2+H \rightarrow NaH+O_2.$$

A reaction with O_3 as follows has also been suggested,

$$Na+O_3 \rightarrow NaO+O_2.$$

The dissociation of the sodium compound and the excitation of resulting Na atom may proceed as follows:

$$NaO+O \rightarrow Na^*+O_2$$
$$NaH+O \rightarrow Na^*+OH.$$

According to Bates and Nicolet [123] the former of the two reactions is possibly endothermic but the latter is certainly exothermic.

It should, however, be pointed out here that from a close examination of the various possible collisional processes which may excite the sodium atom Bates and Nicolet [123] have come to the conclusion that for any of the above processes to be effective, the effective level of emission cannot be much above 70 km. If, therefore, the emitting layer be in the region of F-layer, as measured by Roach and Barbier (see Table II, Sec. 6), one has to seek some excitation process different from those discussed above.

In this connection one may consider the possibility of emission in course of neutralization of the Na^+ ions (which are undoubtedly present in large proportion) [126]. For example, it may be imagined that in the radiative recombination

$$Na^+ + e \rightarrow Na^* + h\nu,$$

the neutral sodium atom is excited to some high energy level of the S, P, D or F state and thence drop to the ^2P state, from which the D-lines are emitted. In such case, emissions of some of the higher members of the sharp or the diffuse series should also be expected in the air-glow spectrum. But, most of these lines may be in the region of the N_2 bands and in the ultraviolet within the Herzberg bands and thus escape detection. It might be mentioned that the presence of λ 3303 has been reported by some workers. Another possibility is ionic recombination analogous to process (I) discussed in sub-section (ii) above. Thus,

$$Na^+ + O^- \rightarrow Na^* + O.$$

Since, ionization potential of Na is 5·12 eV, electron affinity of O is 2·2 eV and excitation potential of Na to the ^2P-state is 2·1 eV, there is an excess of 0·8 eV over exact resonance. If, however, we adopt Vier and Mayer's value [127] of electron affinity of O, namely 3·0 eV [which appears doubtful (see p. 623)], then there may be exact resonance.

A closer examination of the two last named processes shows that both of them are untenable. For radiative recombination an impossibly high concentration of Na^+ ion is necessary. For the mutual recombination process, though the initial concentration is reasonably low, it is not sufficiently large to maintain the observed constancy in the rate of emission throughout the dark hours of the night. There must, therefore, be some process by which the neutral Na atoms produced are being constantly re-ionized during the night. Such night time re-ionization may be imagined to be produced by impact of extra-terrestrial particles.

Alternatively, one may imagine that sodium atoms entering the upper atmospheric regions from interstellar space are being excited by collision with atmospheric particles.

(v) *Meinel bands* (OH).—According to contemporary ideas the OH bands are emitted during, rather than after, the formation of the OH radical [128]. (See Sec. 8.) The most likely process is two body collision between H atom at O_3 molecule as follows:

$$H+O_3 \rightarrow OH^*+O_2.$$

The energy released in the process is just sufficient for the 9th vibrational quantum number level which is the highest observed in the bands. The photon yield for this process at the 65 km. level is found to be about 3×10^5 per cm.^{3} per sec. This is in conformity with the result of photometric measurement according to which there is an emission of 4×10^{10} photons per sq. cm. column per second in the wavelength range λλ 6400–11600 [129]. If the emitting layer is taken as 10 km. thick, then this means that 4×10^4 photons are emitted per cm.^{3} per sec. However, if account is taken of the fact that the emission extends far into the infra-red region of the spectrum, the total emission may be of the order 1×10^5 photons per cm.^{3}

Another possible process of excitation has been suggested by Krasvoski [129*a*]. According to this process large amounts of N_2O is formed as a result of three-body collision in the region where O_2 is dissociated. The N_2O molecule then reacts with atomic oxygen forming nitrogen and excited O_2 molecule. The latter on collision with H atom yields excited OH. We thus have the reactions:

$$\begin{aligned} N_2+O+M &\rightarrow N_2O+M \\ N_2O+O &\rightarrow N_2+O_2^* \\ O_2^*+H &\rightarrow OH^*+O \end{aligned}$$

However, the proposed mechanism may be criticized on the ground that the process can be effective only in the region where O_2 is copiously dissociated, i.e. round 100 km. and above. And this height is much greater than the measured height of the region from which OH bands are emitted (see Table II).

(*b*) Twilight flash

The solar ultraviolet rays besides ionizing and dissociating the upper atmospheric gases must also be expected to produce optical excitation of the atoms and molecules in the same. Radiations due to such excitations cannot obviously be detected during daylight hours. During the morning and evening twilights, however, when the rays of the rising or the sinking sun illuminate the high regions of the atmosphere (see Appendix, Sec. 1) radiation flashes due to such excitations are observed. (It may be mentioned that there has been contemporary attempts to observe 'daylight' air-glow mfro rockets at high altitude. See Chapter XII.) A convenient way of studying the twilight flash is to point the instrument to the eastern or the western sky a few degrees above the horizon and make observations when the sun is rising from below the horizon or sinking below the same.

In the present section we will discuss the *'twilight flash', i.e. considerable* enhancements in intensity—ten to hundred times—that have been observed in the cases of the following lines and bands:

(i) Red oxygen lines.
(ii) OH band λ 6560.
(iii) Yellow sodium lines.
(iv) First negative bands of N_2^+.

(i) *Red oxygen lines.*—Many observations have been made on the twilight enhancement of the red oxygen lines [36, 108, 130, 131] (Fig. 29, Plate 2). The main facts as observed are as follows. The intensity increases 10 to 50 times that of the night time value. The enhancement occurs several hours before morning twilight and several hours after evening twilight and its trace can be observed even when the sun is shining on the atmosphere above 1,300 km. [This, of course, is in accord with the great heights of sunlit auroras as reported by Störmer. See Chap. VIII, Sec. (2*f*)]. The morning and the evening enhancements are not symmetrical, the latter being much larger [36, 130]. According to observations of Cabannes and Garrigue the layer of maximum luminosity lies between 115 and 120 km. [108].

The enhancement of the red oxygen lines is generally ascribed to the resonance effect—solar rays of the wavelengths at λ 6300 exciting the O-atom to the ^1D state. However, the fact that the evening twilight enhancement is much stronger than the corresponding morning one shows that the fact that the upper atmosphere had been bathed in sunlight during the day has a pronounced effect. Thus the enhancement cannot be explained as solely due to resonance.

It is interesting to note that while the red lines are strongly enhanced in twilight, the green line (though stronger in the night air-glow) shows scarcely any enhancement [36]. (A small effect has, however, been noticed by Dufay and Dufay [132]). This difference in the twilight flash is due to the effects of two factors combining to act in the same sense. Firstly, the probability of stimulated transitions of the O-atoms, by absorption of radiation, from the ground state to the ^1S state is very small compared to that to the ^1D state. (See Fig. 17.) Secondly, the intensity of the solar radiation at λ 2972, corresponding to transition to ^1S state, is about 100 times smaller than the intensity of the radiation at λ 6300 corresponding to transitions to the ^1D state. These effects have been studied in detail by Ghosh [116].

(ii) *Sodium D-lines.*—The twilight enhancement effects of this radiation are discussed in Sec. 9.

(iii) *Red* OH *bands* λ 6560.—According to Elvey and Farnsworth [36] the behaviour of this radiation during twilight is very similar to that of the sodium *D*-lines. The increase in intensity during twilight is 8 to 10 times of its night value and occurs as abruptly as in the case of the sodium lines. The enhancement is generally supposed to be a resonance effect.

(iv) *First negative bands of* N_2^+.—These bands are totally absent in the night air-glow but appear with considerable intensity when the upper atmosphere is illuminated with the rays of the sun below the horizon (Fig. 29, Plate II.) We quote below the interesting remarks of Slipher [55] who was the first to make the observation:

'But the negative bands typical of auroral display spectra do not accompany the chief yellow (green) line of the night sky, except when auroral displays are actually present. However, these nitrogen bands, I found, could be photographed in the morning and evening skies if brief exposures were made at the moments when the last and first traces of sunlight touch the high atmosphere. Thus the day, as it were, begins and ends with a sort of auroral flash.'

Observations on the intensity variation of the negative bands during the change of twilight are meagre compared to those for the radiations mentioned above.

According to Dufay and Dufay [132] the flash is of very short duration and occurs when the atmosphere in a layer about 50 km. thick, centred at about 100 km., is illuminated. The same authors also report that the intensity of the flash varies greatly. Sometimes it may be too weak to observe and sometimes it may attain considerable prominence. Dufay and Dufay also observed an interesting effect which is of significance in the theory of the flash. A positive correlation was established between the intensity and the magnetic activity and also between the intensity and the occurrence of auroras in Norway. (The observations were carried out in France.)

The two processes for excitation of the bands that have been suggested are as follows:

(1) $N_2 + h\nu \rightarrow N_2^+ (A') + e$.

(2) $N_2^+ (X') + h\nu \rightarrow N_2^+ (A')$.

Hence, $N_2^+ (A') \rightarrow N_2^+ (X')$ + negative bands. (See Appendix Sec. 2, Fig. 17.)

In the first process, advocated by Saha [133], neutral nitrogen molecules in the ground state are ionized and, at the same time, the ions produced and raised to the $N_2^+(A')$ state (from which the first negative bands are emitted) by absorption of $\lambda < 661$ A.

In the second process advocated by Wulf and Deming [140] it is assumed that $N_2^+(X')$ ions in the ground state are already present in sufficient numbers in the upper atmosphere and are excited by absorption of radiation of appropriate wavelength (λ 3914 for $0 \rightarrow 0$ transition) to the $N_2^+(A')$ state.

According to Bates [134], who has critically examined the two processes, the second is the more likely process.

For the first process, if the sun radiates like a black body (6000°K), the number of available quanta in the wavelength region λ 661 A is found to to be too small. Hence, Saha assumes that there are intense line radiations

near this wavelength range besides the black radiation. However, according to Bates' estimate, the number of photons produced by this process during twilight illumination is too small to account for the observed intensity (10^9 transitions per sec. per sq. cm. column in the line of sight) unless it is assumed that on the occasion of a flash the ionization rate in the F_2-layer is greatly increased. It is also pointed out that on account of the high absorption coefficient of the active radiation λ 661 (see Appendix Sec. 2) the effective shadow of earth for twilight enhancement will be increased by several hundred kilometers. Hence, only the very highest regions of the atmosphere (perhaps above 500 km.) will be illuminated by the active radiation during twilight. This contradicts the observation of Dufay and Dufay that the height of the emitting layer is centred at about 100 km.

For the second process the number of photons that result is given by $(\omega_e/\omega_n)\tau^{-1}$ exp. $(-h\nu/kT)f/$ illuminated N_2^+ ions per second, where ω_e and ω_n are the statistical weights of the excited and the ground states, τ is the radiative life time, f is the dilution factor ($5{\cdot}4\times10^{-6}$), T is the temperature of solar black body (6000°K) [134, 137]. The value of τ according to Bates is of the order 10^{-7} sec. With this value of τ it is shown that the effective emission rate is so large that even if the $N_2^+(X')$ ions form a small fraction of the total ion content the required order of intensity of the flash is obtained.

One important point has, however, to be noted in connection with the calculated value of τ. In the well-known auroral after-glow experiments of Kaplan [138] the negative bands are observed to persist several seconds after the exciting voltage has been cut off (see Sec. 4(*d*)). This means that either the excited $N_2^+(A')$ ions (from which the negative bands are emitted) are actually being produced during the after-glow, or, the radiative lifetime τ of the excited $N_2^+(A')$ ions is to be measured in seconds instead of in tenths of microseconds. If τ has the former value, then the resonance absorption becomes negligible and so also the intensity of the flash spectrum.

8. A NOTE ON HYDROGEN IN THE UPPER ATMOSPHERE

Hydrogen is present in the atmosphere at the ground level in extremely small and doubtful proportion (Chap. XIII, Sec. 1). Further, on account of its extreme lightness it cannot be a permanent constituent of the upper atmosphere as any quantity introduced therein will escape rapidly. No hydrogen lines have been reported in the night air-glow spectrum or in the twilight flash. For these reasons hydrogen in the upper atmosphere had received little attention. However, the discovery by Meinel of strong radiations in the near infra-red as due to vibration-rotation bands of OH (Sec. 4c) has evoked interest in the existence of this gas in the upper atmosphere and on the possible modes of its production.

The various possible processes by which free hydrogen may be produced in the atmosphere may be listed as follows after Bates and Nicolet [135].

$$\text{(i)}\quad H_2O+h\nu\rightarrow OH+H.$$

Of the various hydrogen compounds water vapour is the most abundant in the troposphere, its partial pressure being about 10^{-2} atmosphere. In the stratosphere the fractional volume concentration may be estimated to be 10^{-6}.

The spectral regions responsible for photo-dissociation are $\lambda < 1800$ A and $\lambda < 1340$ A. Some of the radiation in the former region is absorbed by the Runge-Schumann continuum (O_2) and may not reach much below 100 km. But a significant part is not readily attenuated. The latter (including L_α) can reach to about 70 km. level and can produce predissociation. (See Appendix, Sec. *2d*.) It is, however, estimated that the rate for the former radiation is by far the faster.

$$\text{(ii)}\quad CH_4+h\nu \rightarrow CH_3+H.$$

Methane has a continuous source in rotting vegetation and lightning discharges and forms a permanent constituent of the lower atmosphere. Its total amount by volume is estimated to be $2{\cdot}2\times10^{-6}$ of the major gases. However, the fractional concentration is not affected by the negative temperature gradient as in the case of water vapour (due to condensation). Hence, in the stratosphere the methane content might be comparable to the water vapour content.

The photo-dissociation of CH_4 occurs in the same manner as that of H_2O.

$$\text{(iii)}\quad H_2+O \rightarrow OH+O.$$

Molecular hydrogen is estimated to be present in the lower atmosphere in the proportion $0{\cdot}5\times10^{-6}$ by volume of the major gases. As in the case of methane, its fractional concentration is not affected by the negative temperature gradient in the troposphere like that of water vapour. Hence, H_2-content in the stratosphere may also be comparable to that of water vapour.

H_2 is not readily affected by solar radiation; it can, however, react with atomic oxygen producing H and the OH radical as shown above.

The OH produced in reactions (i) and (iii) will produce H by the reaction

$$\text{(iv)}\quad OH+O \rightarrow O_2+H.$$

The energy of activation is known to be low in the reaction. Hence, the rate-coefficient for the reaction must be high.

The hydrogen produced will be lost by the following reactions.

$$H+O_3 \rightarrow OH+O_2.$$

$$H+O+M \rightarrow OH+M.$$

The effectiveness of both these processes for removing H will, however, decrease with height. The first, because of the scarcity of O_3 and the second because of the decrease of the number density of the third body M. Detailed study shows that even for comparatively low regions, the concentration of H will far exceed that of OH. (The authors also consider the

reactions in which other hydrogen compounds like perhydroxyl (HO_2) and hydrogen peroxide (H_2O_2) are produced and the subsequent reactions. Their effects, however, are shown to be small.)

The authors then consider in detail the equilibrium in atmospheric region 60 km. to 100 km. in a pure oxygen atmosphere (containing O, O_2 and O_3) and also in an oxygen-hydrogen atmosphere when bathed in solar radiation. Their investigations show that a thin layer of high H_2 content (concentration $\simeq 10^{10}/\text{cm}^3$.) is produced within 60 to 80 km. region where the temperature falls to a minimum, i.e. near 70 km. The equilibrium in the region is largely controlled by

$$H + OH \rightleftharpoons H_2 + O.$$

According to the authors the hydrogen compounds are completely broken down from above 90 km. (or from even below) and there is a continual escape of hydrogen atoms into interplanetary space [135*a*]. (The concentration of the H-atoms in the neighbourhood of 100 km. is 10^8 to 10^9 per cm^3.) The liberated oxygen atoms, however, remain in the atmosphere and the number that has thus accumulated since geological times is comparable with the number now present. This, in fact, lends support to the view that the photo-dissociation of water vapour and subsequent escape of hydrogen is the origin of oxygen in the atmosphere [136].

9. A NOTE ON SODIUM IN THE UPPER ATMOSPHERE

(*a*) Introduction

The identification of the prominent yellow radiation in the night air-glow spectrum as the sodium *D*-lines (see Sec. 4(*c*)) proves conclusively the presence of sodium atoms in the upper atmosphere.

The intensity of the emission (namely, that corresponding to 2×10^7 to 8×10^7 transitions per sec. per cm.^2 column) is comparable to that of the other line emissions in the night air-glow [57, 94, 123]. The intensity remains fairly constant throughout the night [36] but is subject to considerable fluctuations from night to night. Further, the distribution of intensity is not uniform over the whole sky but is of a patchy character. During the morning and evening twilights the intensity, in the illuminated upper atmosphere, is greatly enhanced—50 to 100 times that in the night air-glow.

The intensity is also subject to seasonal variations, being maximum in winter and minimum in summer.

(*b*) The twilight flash

(i) *Excitation process.*—It is now generally agreed that the sodium twilight flash is simply a case of resonance excitation by λ 5893 [144]. However, there is also the view that the exciting radiation is in the ultraviolet range $\lambda\lambda$ 2000–3000. (The reason for this latter assumption will be discussed presently).

Since the exciting radiation in reaching high atmosphere from the sun below the horizon, has to pass through a considerable thickness of air it is important to enquire into the effect of absorption by the latter. It is obvious that if the exciting radiation be in the ultraviolet range λλ 2000–3000, then it must pass above the ozone layer to avoid absorption by the intense Hartley bands. The resonance radiation λ 5893, on the other hand, is unaffected by this absorption. Nevertheless, some authors [145] assume that this radiation is also affected by ozone absorption—not by the Hartley bands—but by the much feebler Chappuis bands. A little consideration shows, however, that this absorption cannot sufficiently reduce the intensity of λ 5893 to affect the resonance excitation process [126]. For a solar ray incident obliquely through the earth's atmosphere, the maximum value of the ozone mass traversed is approximately 11 cm. and this occurs when the distance of the ray from the earth's surface is 12 km. [146] (see Chap. IV, Sec. 9). The maximum absorption coefficient in the Chappuis bands being ·05 cm.$^{-1}$ (at 6100 A), the maximum reduction in intensity of λ 5893 may be only 40%. If the solar rays pass by the top of the ozone layer (40 to 50 km.), the absorption will be insignificant. There will thus be no significant effect of ozone screening in the region of λ 5893.

(ii) *Height of the region of the twilight flash.*—The region of the twilight flash may be located by noting the height of the moving edge of the earth's shadow (in the line of sight) at which the flash appears (lower limit) and also the height at which it disappears (upper limit). The former gives the height of the lower border and the latter that of the upper border of the region of the flash. The heights so determined obviously depend upon whether one assumes or not the increase in the effective radius of the earth by ozone screening. Thus, Bernard [141]—who was the first to make such measurement—obtained the height of the 'upper limit' as 60 km. *without* taking into account any effect of ozone screening. Hence, he concluded that the bulk of sodium must be below this level. Vegard and Tönsberg [142, 143] on the other hand, admitted the effect of ozone screening and found the value of the upper limit to be 110 km. They also determined the lower limit and, hence, estimated the thickness of the sodium layer in the region of the twilight flash to vary between 8·4 km. and 27 km. with a mean value of 16·2 km. They thus concluded that the sodium producing the flash is contained in a comparatively thin layer between the heights 85 km. and 110 km. (These authors also determined the screening height as 49 km.) The effect of ozone screening was also taken into account by Barbier [131] who, from twilight observation carried out in Haute-Provence (France), estimated the height of the lower limit—the base of the sodium layer—to be at 70 km.

However, in view of the discussion given above any effect of ozone screening on the radiation responsible for the twilight excitation of sodium seems extremely doubtful. If this be so then the base and the top of the sodium layer producing the sodium flash are found to lie approximately at 35 km. and 65 km. respectively.

Note.—The reason why Vegard and Tönsberg assumed the effect of ozone screening is as follows. In their measurements of the height of the 'upper limit' they found that the heights as deduced from observations made when the line of sight was towards a direction near horizon were always higher than those when the line of sight was towards the zenith. This systematic difference, they maintained, could be explained if the exciting radiation was assumed to pass above the ozone layer. It is to be remembered, however, that the height measurements of Vegard and Tönsberg could not have been very accurate. The exposure of the spectrograph for measurements close to the zenith was always much larger than that for measurements near the horizon (2-3 min. near the horizon and 6–10 min. and sometimes 20–30 min. at the zenith). The systematic difference which they observed in the height measurements could possibly have been due to this. In fact, an error of 5 min. would lead to an error of 30 km. in the height measurement.

(*c*) Nocturnal Glow

(i) *Region of the glow.*—Many measurements have been made of the height of the 'layer' emitting the *D*-lines in the night sky by the well-known van Rhijn technique. But, as will be seen from Table II, Sec. 6(*c*), the results obtained are widely divergent amongst themselves. However, in view of the many precautions taken, the results as obtained by Roach and Pettit, namely, 250 km. (in the *F*-region), appears to be more reliable than others.

(ii) *Excitation process.*—This has been discussed in Sec. 7(*a*).

(*d*) Distribution of sodium in the upper atmosphere

From the discussions as given above it would appear that sodium is concentrated in two layers: one in the 35 to 65 km. region responsible for the twilight flash and the other in the *F*-region of the ionosphere responsible for the nocturnal glow.

Alternatively, it may be imagined that the sodium is spread from a height of 35 km. to the *F*-region of the ionosphere. At lower altitudes the sodium exists in the neutral state. At higher altitudes (above 100 km.) all the sodium atoms are ionized.

Resonance excitation of neutral sodium by solar radiation λ 5893 gives the twilight flash of sodium in the lower region.

The sodium 'layer' emitting the night sky *D*-lines is located much higher in the *F*-region of the ionosphere. The layer must consist of ionized atoms, since, according to Bates [149], sodium above 100 km. cannot exist in the neutral state. The *D*-lines in the night air-glow are emitted according to processes suggested in Sec. 7*a* (iv).

An estimate of the total number of sodium atoms in the region of emission of twilight flash may be made from the observed intensity of the twilight emission, and from the flux of solar quanta (taking account of the Fraunhofer absorption) using the known probability of transitions yielding the *D*-lines. The number of sodium atoms per cm.^{2} column is thus estimated to lie between 5×10^8 to 5×10^9 [147, 131]. This constitutes about 10^{-12} of the number of atmospheric particles in the region concerned. (In

view of this minute proportion of sodium the interesting suggestion has been made by Bates [148] that it can be increased by a significant amount by ejecting sodium vapour in this region from a rocket.)

The relative concentration of sodium in the F-region of the ionosphere is much higher than that in the 35 to 65 km. region.

(e) Source of upper atmospheric sodium

Nothing definite is known about the source of the upper atmospheric sodium. It may be of terrestrial origin, but there are strong reasons to believe that at least a part of the sodium comes into the atmosphere from outer space. The various suggestions made regarding the origin are as follows:

Ascending currents of air may carry sodium salt from ocean sprays.

Volcanic dust containing sodium compounds have been known to shoot up to great heights—30 km. [150].

The sodium is carried into the earth's atmosphere by meteorites. (The D-lines have been detected in meteoric spectra. But, it has been argued that the lines may be due to excitation of atmospheric sodium atoms [151].)

The sodium may come from the sun with the solar corpuscular streams which produce auroral and magnetic disturbance phenomena [152].

The sodium really belongs to interplanetary space. It is swept into the earth's atmosphere as the latter moves through this space [145]. (This hypothesis promises an explanation of the seasonal variation of intensity of the D-lines.)

It is not possible in the present state of our knowledge to make any quantitative estimate of the relative importance of the various hypotheses listed above. This is specially because we do not know the rate at which sodium is being removed from the regions from which it emits light.

CHAPTER XI

TEMPERATURE IN UPPER ATMOSPHERE (IONOSPHERIC REGIONS)

1. INTRODUCTION—THE CONCEPT OF TEMPERATURE IN THE HIGH ATMOSPHERE

In Chapters IIIA and IIIB we have discussed how observations on the abnormal sound propagation phenomena and on the luminosity of meteor trails together yield data for constructing the temperature profile of the middle atmospheric regions. To these data have now been added those supplied by direct recordings of pressure variations in this region made in V-2 rocket flights. (See Chap. XII.) We thus have now a fairly accurate knowledge of the temperature distribution up to about 100 km. In the present chapter we shall make an attempt to estimate the temperature distribution round and above this level from a number of upper atmospheric observations and measurements. It will be seen that these estimates, though arrived at by diverse methods, all agree amongst themselves, at least qualitatively to point to the existence of a region of rising temperature above the 100 km. level, the temperature attaining a value of the order 1500°K to 2500°K in the region of the F_2-layer of the ionosphere. Theoretical considerations also lead to the existence of a similar high temperature in these regions. The rise is attributed to the degradation of the energy of extreme solar ultraviolet radiation (both black-body radiation and intense line radiations during solar flares) to heat energy by absorption by the gases in the highest atmospheric regions.

In what follows we shall first discuss briefly if the concept of temperature is applicable to the tenuous atmosphere in the high regions under consideration.

If a gas is in thermal equilibrium, then the gas temperature is equivalent to the so-called kinetic temperature defined by the relation

$$\bar{v} = \sqrt{8kT/\pi m},$$

where $\bar{v}$ is the average velocity and the velocity distribution is Maxwellian.

Now, if the concentration of the particles be low—as in the high atmosphere in the ionospheric regions—one may be justified in the apprehension that the collisions will be so few and far between that there will be large deviations from the average and that, as such, average velocity and kinetic temperature will cease to have any meaning. However, closer scrutiny shows that such apprehension is groundless. We have the relation

$$Q = (2/3N)^{\frac{1}{2}},$$

where Q is the ratio between the standard deviation of energies and the average deviation (N is number density of the particles). For a gas under

ordinary temperature and pressure conditions N is of the order $10^{19}/\text{cm.}^3$ and hence Q is of the order 10^{-10}. For the attenuated gas at 300 km. height, N is of the order $10^9/\text{cm.}^3$ and Q is about 10^{-5}. This value of Q, though many orders higher than that for the gas under ordinary pressure and temperature conditions, is still very small. There is therefore thermal equilibrium in the high atmospheric regions under consideration and, hence, we can speak of kinetic temperature of such regions.

It may be noted, in this connection, that according to modern ideas, even the sparsely populated interstellar medium (population 1 to 10 particles per cm.^{3}) can, for many purposes, be treated as perfect gas with complete kinetic equilibrium between the different particles [1].

We can also assure ourselves that there is sufficient molecular concentration in the high regions for the gas kinetic laws to hold from the simple consideration that the molecules there are prevented from escaping by collision with molecules above (see Chapter I, Sec. 5*c*). Under such condition, the collisions must be so frequent that one can speak of an average velocity of the particles and, temperature as defined by the relation as given above, will have a meaning.

There is another point of importance which needs consideration before we proceed to our discussion on the estimation of upper atmospheric temperature. In the section to follow it will be seen that the most important methods of temperature estimation are those utilizing results of ionospheric measurement. In the interpretation of these results the energy state of the electrons is involved. It is therefore pertinent to enquire if the energy distribution amongst the electrons is subject to the law of equipartition, i.e., if the electron temperature is the same as the gas kinetic temperature. The reason of this query is that the electrons which are produced by photoelectric action (or, by solar corpuscle bombardment) will, in general, possess initially energy far in excess of that corresponding to the gas temperature. And, it is conceivable that on account of their small mass the electrons may not get rid of this excess energy by collisions. However, simple calculation shows that for all reasonable excess value of energy, the electron temperature will settle down to the equivalent gas kinetic temperature in course of only a few seconds. Thus, if an electron is liberated with an energy of 10 eV energy in an atmosphere of mean molecular weight 20 and temperature 300°K, then the electron energy will come down to that corresponding to the gas temperature in only about 25 sec. in the F-region and in a smaller time in the E-region [2].

2. ESTIMATION OF TEMPERATURE AND ITS HEIGHT DISTRIBUTION

The temperature and its height distribution in the uppermost regions of the atmosphere may be estimated from the following:

(*a*) Escape of helium from terrestrial atmosphere.

(*b*) Extension of the atmosphere to great heights.

(*c*) Height-luminosity distribution in auroras.

(*d*) Ionospheric measurements:

(i) Collisional frequency of electrons.

(ii) Scale height values.

(iii) Recombination coefficient values.

(iv) Diurnal variation curve of electron concentration (deviations from the $\sqrt{\cos \chi}$ law).

(*e*) Width of emission lines from the night air-glow.

(*a*) Escape of helium

The problem of the escape of helium has been discussed in Chapter I, Sec. 5(*b*), where it has been shown that helium would escape in appreciable quantity from the terrestrial atmosphere if the helium atoms, in the region of escape in the high atmosphere (600 to 800 km.) have thermal velocities corresponding to temperature of the order 1000°K.

(*b*) Extension of the atmosphere to great heights

It has been found that twilight flash of the red oxygen lines can be traced up to the height of 1,300 km. (Chapter X, Sec. 7.) Also auroral streamers, particularly those in the sunlit atmosphere, have been observed up to heights as great as 1000 km. These facts show that atmosphere, as such, must exist up to these heights, and, as shown below, this is only possible if there is a rising gradient of temperature in regions above 100 km. The molecular concentration prevailing at such heights of the atmosphere can easily be calculated (on the assumption of thorough mixing) if the density and temperature at some datum level, as also the distribution of the latter in the region above be known. The datum level may conveniently be taken as 100 km. where the temperature and molecular concentration are known to a fair degree of approximation from radio measurements (300°K and 3×10^{13} per cm.^{3} respectively). Now if a constant temperature be assumed to prevail in the region above 100 km. then the concentration at 400 km. is found to be only three molecules per cm.^{3} Such concentration is totally inadequate to produce the observed intensity of the twilight flash or of the auroral streamers. Further, the concentration is wholly incompatible with the measured electron concentration in this region, the average value of which is of the order 10^6/cm.^{3} If, however, a rising temperature of 4°K per km. is assumed (which makes the temperature at 300 km., 1100°K) then the concentration at 400 km. becomes 10^7 molecules/cm.^{3} If account is taken of the fact that oxygen in the upper atmosphere exists mostly in the atomic state then the density at 400 km. level is found to be still higher, viz., 10^9 atoms/cm.^{3} Such concentration is quite adequate to produce the observed twilight and auroral luminosity effects.

An interesting point regarding the estimation of the temperature gradient from the twilight flash of the red-oxygen lines mentioned above may be noted here. Bates [3] has shown that during this twilight flash the

number of photons emitted per cm.^{3} per sec. is 2·3 at 300 km. and 0·35 at 700 km. One is tempted from this to assume that the populations of normal oxygen atoms at the respective heights are proportional to these numbers and hence conclude that the number density of O-atoms has decreased by a factor of 6·6 only in going from 300 km. level to 700 km. level. This decrease of density with height is compatible only with a temperature of 3600°K for the region. The conclusion is, however, not justified, because, at the lower level an appreciable fraction of the excited ^1D atoms will revert to the normal state by collisions and not by radiation of the red quantum of light. Hence, the total number of ^1D atoms de-excited per second per cm.^{3} is not proportional to the intensity of the emitted line at all heights. The fall in molecular concentration with height is, therefore, much more rapid than that calculated without taking this factor into consideration.

(*c*) Height-luminosity distribution in auroras

In Chapter VIII, Sec. 2*e*, we have seen how the rate of absorption of electron rays along their trajectories determines the distribution of luminosity in auroral streamers. The rate of electron absorption at any height obviously depends upon the number of density of the absorbing particles at that height. And, since the logarithm of the number density is inversely proportional to the scale height of the region, it is clear that the measurement of the gradient of luminosity enables one to estimate the value of the scale height in the region of observation. Harang [4, 5], by thus utilizing the height-luminosity functions of auroras, has deduced scale heights in the region 100–350 km. In the measurements made by Harang, the computation of H for regions below 185 km., was made from data on auroral arcs and draperies; at 185 km. from auroral rays; at and above 270 km. from examination of sunlit rays. According to Harang the scale heights as deduced by him give the lower limits of the same, because, the electron rays have been assumed to proceed along straight paths and not along spiral paths round the lines of force as they actually do (Chapter VIII, Sec. 5*c*).

(*d*) Ionospheric measurements

(i) *Collisional frequency of electrons.*—In Chapter VI, Sec. 7, we have mentioned how frequency of collisions of electrons with different kinds of atmospheric particles may be determined from ionospheric data. Since the collisional frequency is a function of temperature, the latter may be estimated from the measured values of the former. The collisional frequency measured from ionospheric parameters in the region under consideration is the sum of two collision frequencies: One between electrons and neutral particles, and the other between electrons and positive ions. (There may also be collisions between electrons and electrons; but, for obvious reason the frequency of the same is negligible compared to the two mentioned above.)

In the case of collisions of electrons with neutral particles we have, for the collision frequency, the expression (see also Appendix, Sec. 5),

$$\nu = 2\sigma^2 P/(3m_e kT)^{\frac{1}{2}},$$

where P is the partial pressure of atoms and molecules and σ is the collisional cross-section.

For collisions between electrons and positive ions the expression as given by Mazumdar [6] is,

$$\nu' = (4\pi e^4/_3)/(2\pi m_e k^3)^{\frac{1}{2}}\,(n/T^{\frac{3}{2}})\,[\log\,(T\bar{\lambda}')^2 - B]$$

where $B = 24{\cdot}58$ and $\bar{\lambda}'$ is the mean ionic free path, n, number of ions per cm.^{3}

Utilizing these two formulae and taking into account the most probable cross-sections of the neutral particles—O, N, N_2 [special attention being given to the cross-section of O which depends on electron energy as well as on atomic field forces (see Chapter VI, Sec. 7)], Gerson [2] has determined the values of probable temperatures at 300 and 400 km. (Table I, p. 556).

It is to be noted that in the above formulae pressure, at the level under consideration, is involved and this itself depends upon the temperature. However, if a linear temperature-height variation is assumed, then with the help of Eq. (3) in Chapter I, the pressure term can be eliminated.

(ii) *Scale height measurements.*—It has been shown in Chap. VI, Sec. 6, that the thickness of a simple Chapman layer is equal to $4H$, where H is the scale height of the active constituent in the region where the layer is formed. The approximate average values of H for the different ionospheric regions (on the assumption that they are all of the simple Chapman type) are as shown below:

Height (km.)	70	100	200	200	250	300
$H = kT/mg$	6	10	35 Winter noon	50 Summer noon	50 Winter noon	70 Summer noon

A better estimate of H on the assumption of a parabolic layer may be made from Eq. (75), Chapter VI, Sec. 6.

The parabolic layer, however, is only a rough approximation to the Chapman distribution. One may use, instead of the parabolic approximation, any of the various higher approximations [7, 8, 9] to Chapman layer that have been developed. The values obtained with such approximations are somewhat different from those obtained on the parabolic assumption, being generally higher.

It must also be remembered that the Chapman type of distribution is produced only under highly idealized conditions, namely, when the temperature and the recombination coefficient remain constant in height. Under

actual conditions, however, both these parameters vary and, as such, there is considerable distortion in the Chapman distribution. As a matter of fact, increase of temperature and decrease of recombination coefficient with height both give values of H higher than the true value. As such, the values of the temperature, as deduced from a consideration of the values of H, indicate only the upper limit of the true value.

For the E-layer the matter is further complicated by the fact that it is formed in the region of transition of O_2 to O and, hence, there is decrease in the mean molecular mass m with height. In this connection mention may be made of the recent work of Pfister [10], according to which if the magnetic field of the earth is taken into account the value of the scale height in the E-region is lowered by about 20%.

Subject to the above limitations, the value of H as also its variations indicate that the upper limit of the temperature in the region of the F-layer (200 km.–300 km.) varies between 600°K and 2500°K.

(iii) *Recombination coefficient values.*—The observed variation with height of the coefficient of recombination (α) of ions and electrons provides another ionospheric data for estimating the temperature. The relation between α and T is given by the formula of Thomson and Thomson [11],

$$\alpha = cT^{-\eta}$$

where c and η are constants. Hence, if the value of the exponential η is known, the variation of temperature with height may be estimated from the variation of α with the same. Unfortunately different workers have assumed different values of η. Thus, Seaton [12] has assumed $\eta = 3$ and has obtained temperatures ranging from about 1000°K to 100°K for the F_2-layer, 1700°K to 50°K for the F_1-layer and 570°K to 70°K for the E-layer, depending upon latitude, season and time of day and night. Baral and A. P. Mitra [13] have favoured the value 1 for the E-layer. Yonezaswa [14] on the other hand, working in Japan favours the value $\frac{1}{2}$ for η, but his calculations give only relative temperature changes in the F_2-layer. These data indicate the existence of a high temperature above 100 km. level, though, in view of the uncertainty in the value of η, the actual value of the temperature as deduced is to be taken with reserve.

(iv) *Diurnal variation of electron concentration (deviations from $\sqrt{\cos\chi}$ law).*—According to Chapman the maximum number density of electrons for any of the ionized regions varies as $(\cos\chi)^{\frac{1}{2}}$, where χ is the zenith angle of the sun. (Chapter VI, Sec. 10.) It is generally believed that this type of diurnal variation holds accurately at least for the E and F_1 regions. Careful analysis [15, 16] has shown, however, that the exponent of $\cos\chi$ actually varies from 0·50 to 0·80. This increase in the value of the exponent can be explained if it is assumed that there is a rising temperature gradient. In fact, for a temperature gradient γ°/km., it can be shown that the exponent is equal to $1+(\gamma k/mg)/2$. This gives, for conditions over Calcutta, a temperature gradient of about 4°/km. round 120 km. height.

(e) Width of emission lines from the night air-glow

It is possible, at least in principle, to determine the kinetic temperature of the regions of night air-glow (Chap. X) from measurements of the width of the emitted lines. One of the principal causes of the broadening of the spectral lines is the Doppler shift due to thermal motions of the emitting atoms; the line 'width' $\Delta\lambda$ is related to the kinetic temperature by the equation

$$\Delta\lambda = 7{\cdot}2\times 10^{-7}\lambda\sqrt{T/m}.$$

Hence, from a measurement of $\Delta\lambda$ the kinetic temperature of the emitting source may be determined. Such width measurements may be carried out with the help of the Fabry-Perot type of interferometer. Unfortunately, on account of the faintness of the light, it is difficult to obtain the high resolving power (by increasing the separation of the interferometer plates) necessary for the purpose. There is always considerable uncertainty in estimating the sharpness of the fringes, i.e., in estimating the minimum contrast needed to make the rings appear sharp. The accuracy of such width measurements has, therefore, been questioned by Spitzer [23]. Interferometer measurements of line widths are thus, not very helpful in determining the temperature. They give, at best, a possible upper limit of the kinetic temperature of the emitting source. Nevertheless we may note here the results of some of the interference measurements of line width.

Thus, Babcock [17] from his interferometer measurements of the green oxygen line concluded that its width was not greater than 0·035 A. Hence, applying the formula given above we may deduce that the kinetic temperature of the emitting oxygen atoms was not higher than 1,200°K.

Similar interferometer measurements have been carried out by Vegard and his collaborators [18]. But, the only significant result obtained by these authors is that the sharpness of the interference rings (i.e., the temperature) does not change with height.

Attempts have also been made to estimate the kinetic temperature from intensity distribution within the nitrogen bands in aurora [19], as also in the night air-glow [20, 21]. The results obtained are, however, in sharp contrast to those mentioned above. For the case of auroras a low temperature of the order 220°K has been obtained for all heights up to 800 km. For the night air-glow, temperatures between 150°K and 230°K were obtained for the region at 90 km. Such 'rotational' temperatures cannot be expected to represent the true kinetic temperature because, the intensity distribution within the bands depends upon the mode of excitation. (For example, in the case of auroras the bands are excited by electronic impacts or by impacts with other charged particles.) Hence, until the relationship between the kinetic temperature and the rotational temperature is better understood one is not justified in making any deduction regarding the kinetic temperature from measurements of intensity distribution within the bands.

(f) Concluding remarks

The values of temperature as deduced above by diverse methods all agree amongst themselves at least qualitatively, to point to existence of a rising temperature gradient above 100 km. level. Gerson [2] has assessed the relative accuracies of these values and has come to the conclusion that those deduced from the auroral height-luminosity curves, from the collisional frequency data and from electron concentration (i.e. ionospheric scale heights) represent most closely the actual temperature distribution. Table I, as prepared by Gerson, gives, for the region 100–400 km., temperatures as calculated by these three methods.

TABLE I

Temperature distribution in the ionospheric regions

Altitude (km.)	Upper Atmospheric Temperature (°K) deduced from		
	Collisional frequencies	Auroral scale heights	Electron concentration
100	300	219	
150	825	531	
200	1350	1580	
250	2175	2073	
300	2400	2455	
330	....	....	1530–2680
350	3225	2704	....
400	3450	....	....
500	....	....	2200–3910

Temperatures higher than those given in the Table are not likely. As a matter of fact, it may be concluded, at least qualitatively, that the maximum temperature is attained somewhat above 400 km. Beyond this height the temperature probably decreases to merge with the temperature of interstellar space.

3. THEORETICAL CONSIDERATIONS

The experimental evidences described above, though more or less qualitative, leave little doubt that there is a rising temperature gradient in the upper atmospheric region above 100 km. and that the temperature at the height of the F_2-region of the ionosphere is some thousand degrees absolute.

Conclusion regarding high temperature is also arrived at from the following simple theoretical considerations.

The whole of the extreme ultraviolet radiation of the sun is absorbed in the upper regions of the atmosphere and fail to reach the surface of the earth. The portion of the spectrum between λ 2100 and λ 2900 is absorbed

by ozone in the middle atmosphere. The rest of the spectral energy below λ 2100 is undoubtedly absorbed by the constituent gases of the upper atmosphere, namely, by molecular nitrogen and by molecular and atomic oxygen. The primary effect of the absorption might be dissociation, ionization or excitation of the constituent gas particles; but, like all other forms of energy, the absorbed energy must ultimately degrade into thermal energy of molecular agitation causing a rise of temperature.

Theoretical calculations on the temperature rise have been made by several authors by making plausible assumptions regarding the gain of heat energy by absorption of solar ultraviolet light and its loss by radiation and conduction.

According to Godfrey and Price [22] the equilibrium temperature due to absorption of radiation $\lambda < 1450$ A by oxygen molecules (which have not been dissociated into atoms) in the upper atmosphere may be as high as 3300°K. If it is assumed that water vapour (which is a good radiator of heat) is also present in this region then the temperature attained may be lower, but still of the order 1200°K.

According to Spitzer [23] the temperature is profoundly influenced by conduction. In the region where oxygen is fully dissociated, the rates of energy absorption and radiation per cm.^{3} are both very substantially reduced. If the sun is considered to be radiating like a black body in the extreme ultraviolet, and the absorption is in the wavelength region $\lambda < 910$ A then the temperature is determined by conduction up from the level at which O_2 is dissociated. The expected temperature has not been determined theoretically but it is much less than 1500°K which, according to Spitzer, is necessary for the escape of helium. If, however, the sun be radiating much in excess of black body radiation in this region of wavelengths (intense line radiation for example) then radiation determines the temperature. For an excess by a factor 10^2 (or more) there will be radiative equilibrium up to considerable heights. In this case the temperature is determined by a balance between photoelectric absorption of energy in the ultraviolet and dissipation of energy by excitation of oxygen atoms by electronic impacts. With large excess in the solar ultraviolet radiation the temperature in the region of the F_2-layer will be between 1800°K and 3000°K. Spitzer is of opinion that the average temperature in the region of 300 km. height is more nearly 500°K, being determined by conduction from lower layers. But during the brief intervals of ultraviolet solar flare, the temperature increases to a high value (exceeding 2000°K) allowing helium to escape.

The problem of radiative equilibrium in the higher regions of the atmosphere has also been examined by Wooley [24]. According to Wooley the high temperature arises out of a heavy ultraviolet absorption and a relatively weak infra-red emission. It is suggested that at heights below 250 km., O_2 is the principal absorbing agent; at greater heights the absorption is due mainly to O. Further, at heights of 100 km., water vapour is the principal radiator; at greater heights (250 km.) water vapour is absent and the temperature is effectively controlled by negative ions.

Bates [25] has considered thermal equilibrium in the *F*-region by considering on the one hand the rate of energy gained from ionizing photons and on the other the various loss processes. It is found that the most important loss process is conduction and that next to it is probably emission from atomic oxygen. According to the author these loss processes are so effective that the energy source appears to be inadequate to maintain the high temperature which the region is usually believed to possess. The tentative suggestion is made that heat may be supplied to the upper atmosphere mainly by non-observed ionization.

CHAPTER XII

ROCKET EXPLORATION OF THE UPPER ATMOSPHERE.*

1. INTRODUCTION

In Chapter I we have mentioned how balloons carrying measuring instruments and apparatus are used for direct exploration of the tropospheric and lower stratospheric regions. The highest altitude reached by such balloons (unmanned) is about 30 km. For the higher regions, rockets of the type invented in Germany, during World War II, as also lighter and smaller rockets specially designed for upper atmosphere research, have been successfully employed. The maximum height reached by such rockets has been 180 km., and useful scientific data have been recorded up to the height of 120 km. The rocket exploration programmes now include the study of almost all the physical aspects of upper atmosphere, e.g. composition, temperature, pressure, humidity, ionization, day air-glow, upper atmospheric electric current systems and solar ultraviolet light absorption.

Most, if not all of the contemporary rocket explorations are being conducted in the U.S.A., the country in which the study was first initiated. The earlier explorations (beginning in 1946) were made with captured German V-2 rockets after reconditioning as necessary. Later, a new model, 'Aerobee', has been developed and is being successfully used.

It is worthwhile putting on record how the upper atmosphere research programme—specially with reference to the German V-2 rockets—was first organized in the U.S.A.

The army Ordnance Department of the U.S.A. announced towards the end of 1945 that they possessed a number of captured German V-2 rockets and that these would be fired to increase the knowledge of rocket techniques and ballistics. At the same time a number of scientific organizations were invited to devise scientific instruments which could be carried by the rockets to record upper atmospheric data. The following participated in this programme:

Naval Research Laboratory (N.R.L.),
Applied Physics Laboratory of Johns Hopkins University (A.P.L.),
Army Air Forces (through the Watson Laboratories of the Air Material Command (A.M.C.) and the University of Michigan),
Army Signal Corps (Evans Signal Laboratory),
Aberdeen Ballistics Research Laboratory.

---

* The materials for this Chapter, specially those for the V-2 rocket, have been gathered largely from the Reports of the Naval Research Laboratory, U.S.A. (Nos. R-2955, 3120, 3030, 3171, 3083). The articles in the publication *The Atmospheres of the Earth and Planets* (University of Chicago, Editor: G. P. Kuiper), under the Chapter heading 'Upper Atmosphere Studied from Rockets,' by J. L. Greenstein, H. E. Clearman and E. Durand have also been very helpful.

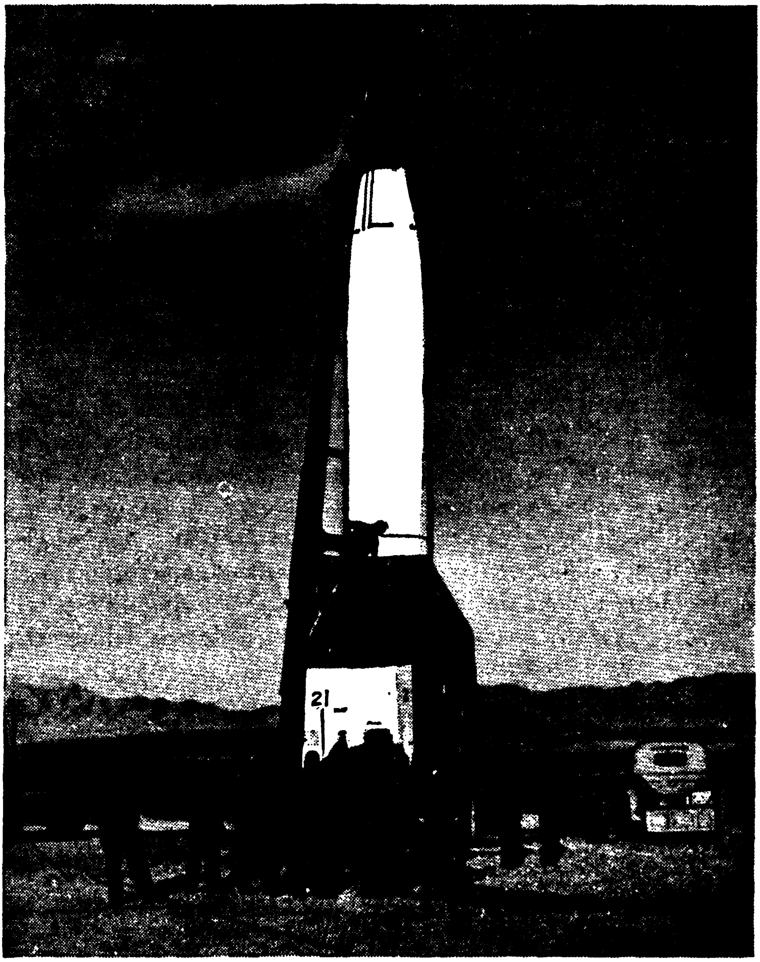


FIG. 1. The V-2 rocket in launching position (Official United States Naval Photograph).

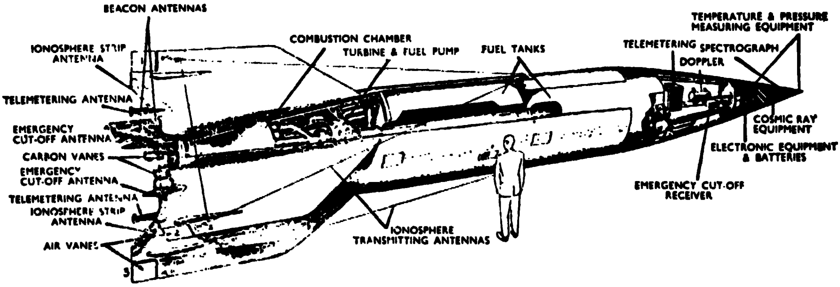


FIG. 2. The V-2 equipped for upper atmosphere study.

Besides the above the Harvard, the Princeton and the Ohio State University groups, the National Bureau of Standards, the California Institute of Technology and the American Meteorological Society co-operated in the programme. The General Electric Company assisted the staff of the White Sands Proving Grounds, New Mexico (which is the rocket testing site of the U.S.A.) in the assembly and firing of the rockets. A V-2 Panel consisting of members drawn from most of the above institutions and agencies co-ordinated all the work. In the first phase of the programme extending up to 1949 more than 50 rockets were made available with salvaged parts and with American replacements.

2. THE ROCKET

(a) Introduction

The rocket is the simplest and the most efficient of all heat engines. It can operate in space devoid of air as it carries its own fuel and oxidizer—alcohol and liquid oxygen. The velocity (V) attained by a rocket, in the absence of any other force, is given by

$$V = v \log_e R$$

where v is the velocity of the jet or exhaust and R is the so-called mass ratio, i.e. the ratio between the initial weight of the rocket and the weight after the fuel is burnt out. For the case of a rocket fired from the earth, the effects of g and of the air-resistance have to be taken into account. For vertical firing we have,

$$V = v \log_e R - gt - F(V, h)$$

where t is the time of combustion and $F(V, h)$ depends in a complicated manner on the velocity, height and shape of the rocket. It may be pointed out that the reduction in the final velocity due to the last term is not more than 10 to 15 per cent, while the reduction due to g may be kept small by reducing the time of combustion.

FIG. 3. The V-2 takes to the air.

(5) The V-2

The reconditioned V-2 rocket was 14·32 m. (47 ft.) long and 1·65 m. (5 ft. 5 inches) in diameter (see Figs. 1 and 2). Fully loaded, it weighed about 14,000 kg. Of this, the weight of alcohol and liquid oxygen was 9,500 kg. Besides, it could carry a payload of about 900 kg. up to about 160 km. altitude. The alcohol and liquid oxygen were introduced into a combustion chamber at a high pressure where the mixture was burnt. They were consumed during the first minute. The combustion occurred near 2,000°C. and the gas jet developed a thrust of 28,500 kg. The maximum acceleration (six times g) was attained soon after take off and just before the fuel is burnt out. The terminal speed attained at the instant of burning was 1·5 km./sec. Graphite steering vanes in the jet controlled the flight in the initial stage till the fuel was exhausted. (See Fig. 3.) Stabilization in the azimuth was maintained by fixed gyroscopes. According to a predetermined schedule, the gyroscope controlled the tilt of the fins. The vanes deflected the direction of the exhaust stream and thus served to keep the rocket on a predetermined course. For upper atmosphere exploration the vanes were adjusted to give a gradual tilt during the powered portion of the flight, so that during the flight the rocket was carried about 50 km. horizontally from the site of launching. After the fuel had been burnt up, the flight was necessarily uncontrolled, and the rocket soared only under the decelerating action of gravity. For this uncontrolled flight the rocket had usually a small angular momentum which was unpredictable. This caused the rocket to roll slowly about its principal axis (once in a few seconds) and at the same time the axis yawed in space.

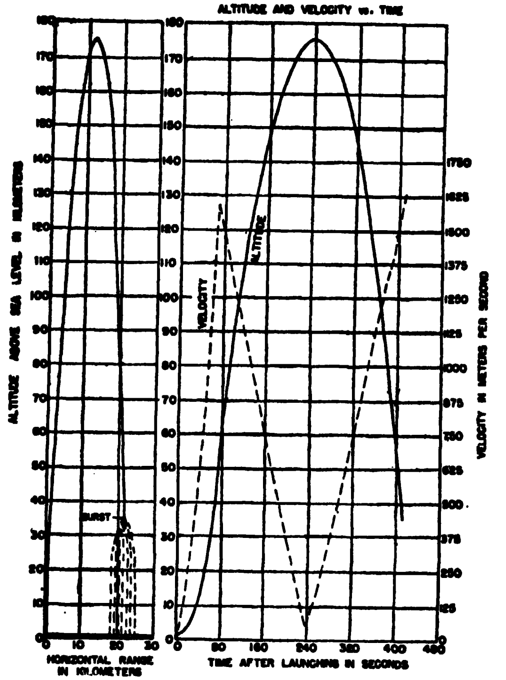


FIG. 4. The trajectory of the V-2 fired on October 10, 1946. The trajectory information was deduced from experimental data.

An average successful flight lasted for about 6 minutes, the average height reached being 120 km. The rocket spent a large fraction of its total time of flight in the upper part of the trajectory. Thus, for a flight lasting 450 seconds (altitude reached 170 km.), 270 seconds was spent above 80 km. (Fig. 4 shows the trajectory of a 161 km. flight.) The landing speed

was very high, about 1 km./s. The impacts created craters about 25 m. wide and disintegrated the rocket.

(c) The Aerobee

The V-2, designed as it was for war purposes, was necessarily elaborate and contained many complicated features not essential for upper atmospheric study. A simpler, smaller and less costly rocket—the so-called Aerobee—has, therefore, been developed and successfully used. The programme of development of Aerobee was initiated by the Applied Physics Laboratory of the Johns Hopkins University, supported by the U.S.A. Navy Bureau of Ordnance with the assistance of the Office of Naval Research [1]. The actual construction, under the technical direction of this Laboratory, was undertaken by the Aerojet Engineering Corporation and the Douglas Aircraft Company. The immediate engineering basis for the rocket was provided by some prior work, on a smaller rocket, carried out by the Jet Propulsion Laboratory of the California Institute of Technology.

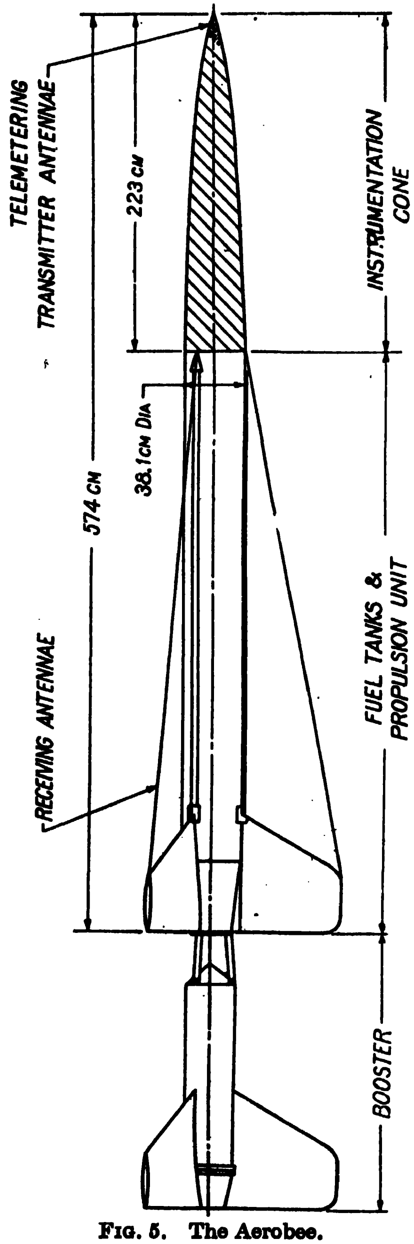


FIG. 5. The Aerobee.

Fig. 5 shows the outline of the Aerobee. It is constructed largely of non-magnetic materials such as aluminium and stainless steel. The diameter is 38·1 cms. (1 ft. 3 inches) and the length 5·74 m. (18 ft. 10 inches); the nose-cone is 2·23 m. (7 ft. 4 inches) long. The nose-cone is pressure-tight and contains the payload of the observational equipment and instruments. The propulsion of the rocket is carried out in two stages. In the first stage a 'booster', using solid fuel, brings the velocity up to about 300 m. per second. The booster then falls away and the sustaining motor (note its fuel tanks in Fig. 5) propels the rocket for about 45 seconds when the fuel is exhausted. At this instant, if the launching had been nearly vertical, the

height attained is 29 km. and the velocity about 1·25 km./s. With a payload of 68 kg. the maximum altitude reached is about 112 km. Unlike the V-2, the Aerobee does not carry any guidance or control equipment. Its stability in flight is attained by the three tail-fins fixed in position and by the proper location of its centre of mass. The trajectory, which the rocket is desired to follow, is controlled by suitably adjusting the tilt of the long launching tower 42·67 m. (140 ft.) in accordance with the local wind data obtained from sounding balloon observations. The appropriate windage theory for this purpose has been developed by the Ballistic Research Laboratories of the Aberdeen Proving Grounds and by the Applied Physics Laboratory of the Johns Hopkins University. The actual firings are conducted by the Naval Unit, U.S.A.; the operation of the telemetering equipment and the necessary field work and the reduction of data are carried out with the assistance of the Physical Science Laboratory of the New Mexico College of Agriculture and Mechanic Arts.

It may be mentioned here that a modified version of the Aerobee—the Navy Viking—is also being developed and has already undergone preliminary tests. Unlike the V-2 and the Aerobee it is stabilized during its free fall—the stabilizing being accomplished by means of air jets [2]. This is a great advantage as the aspect of the rocket (say with respect to the sun) remains unchanged as a result of the stabilization and makes easier the interpretation to be put on the results of pressure, cosmic ray and ionosphere measurements.

These types of rockets, owing to their simpler construction and comparative inexpensiveness, are expected to be widely used in upper atmospheric investigation.

3. INSTRUMENTATION AND DATA RECOVERY

(a) Introduction

The conditions under which scientific observations have to be carried out in the rocket in course of a flight are very exacting. Thus, the pressure varies from 665 mm. (at the elevation of the Proving Ground) to 10^{-4} mm.; the temperature varies from $+60°$ to $-80°$C.; the skin of the rocket is heated to about 200°C.; all high-voltage apparatus have to be kept in pressure-tight spaces. It is, therefore, no wonder that many of the flights (about half of the total) are unsuccessful not only on account of the poor performance of the rocket, but also because of the failure of the apparatus and instruments carried to work under the severe stresses of the flight.

It is essential to track the rocket detailing where it is at all the time. This involves many types of radio, radar and optical methods. (The work involved in this was carried out by the Ballistic Research Laboratory of the Aberdeen Proving Ground.)

(b) Instrumentation

The instruments and apparatus as may be carried by the rocket are as follows: gauges for measurement of air-pressure (different types for

different pressure ranges; *vide infra*), specially designed spectrographs for photographing the ultraviolet solar spectrum, radio transmitters for ionospheric studies, apparatus for collecting air samples, cameras for photographing the earth's surface (to determine at all times during the flight the aspect of the rocket), cosmic rays measuring apparatus, magnetometer to measure the total magnetic field of the earth and, finally, telemetering equipment for transmitting to the ground by radio the various instrumental 'readings', as also the performance of the rocket in course of its flight. These instruments and apparatus are suitably mounted in the different parts of the rocket, the nose-cone carrying most of them. In the reconditioned V-2 the nose, which originally contained the explosive, was replaced by a war-head (Fig. 6).

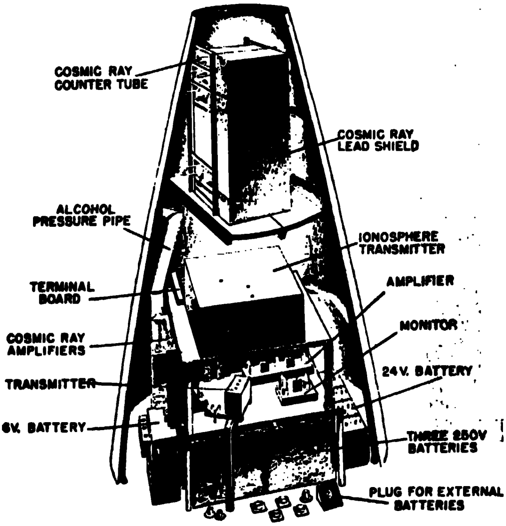


FIG. 6. Instrumentation in the war-head of a V-2.

(*c*) Data Recovery

Two methods are employed for recovering the data at the ground. One is to record the data (usually on films) and recover them after the rocket has crashed to the ground. The other is to telemeter, i.e. transmit to the ground by radio, the 'readings' of the instruments during the flight.

Destroying the stream-lining.—In the first there is the unavoidable risk of damaging the instruments and apparatus and their records due to the impact on landing of the rocket with high velocity. This is avoided by blowing off, when the rocket is at a height of about 50 km., by TNT the war-head. This destroys the stream lining of the rocket and the rocket structure is torn apart by the aerodynamical forces as it enters the denser atmosphere below. The velocities of the small-size component parts are greatly reduced by the large air-resistance and the landing speed is only of the order of 0·1 km./s. (as compared to 1 km./s. for the entire rocket). The

instruments and apparatus recovered by this method are, not infrequently, in such good state that they are used again.

Special parachutes have also been devised and successfully used to land equipment ejected from the war-head at a small speed.

It is also necessary to locate the impacts of the rockets. For V-2, this is easily done with the help of the optical and radar tracking data supplemented by search from air. For the Aerobee, a system of sound ranging, on lines similar to that employed in World War I for locating large gun emplacements, has been developed.

Telemetering.—Complete system for telemetering to the ground the rocket data when it is in flight has been successfully developed by the Naval Research Laboratory. The system supplies, on the one hand, the upper atmospheric data as indicated by the various instruments and, on the other, information about the performance of the rocket when in flight. The physical data of the upper atmosphere supplied are temperature and pressure at various altitudes, different characteristics of the primary cosmic radiation and properties of ionosphere. The rocket performance data telemetered are such quantities as speed, acceleration and altitude, skin temperature at various critical points, and motion of control fins. These data help to locate the cause of any malfunctioning of the rocket during flight, so that future improvements on rocket construction can be made.

The basic method of data transmission from the rocket is to convert data voltages into time intervals and is known as the *pulse-time-modulation* system. The use of pulse system has the advantage of delivering a higher peak power with a lower average power consumption and of occupying less space than a comparable continuous carrier system. The frequency of operation of the system is about 1,000 Mc./s. The frequency is chosen because this is high enough to penetrate the ionosphere, and at the same time is clear of other frequencies used in the flight.

FIG. 7. A telemetering receiving antenna installed at White Sands Proving Ground, New Mexico, for signals telemetered from V-2.

The telemetering system as used in the U.S.A. is a multichannel system. The voltage presented to the input of any of the channels determines the spacing between two adjacent pulses defining the channel. Channel 1 determines the spacing between the first and the second pulses, channel 2 that between the second and the third, and so on. Thus, for a complete sampling of the data in V-2, 23 in number, 24 pulses were required. Two hundred such samplings are made each second, a master keyer initiating the sampling process at a uniform rate. The distinction between different groups is made by

allowing for a sufficiently long time between the initial pulses of each group. This makes the interval between the last pulse of one group and the first of the next very much longer than any of the measuring intervals.

The data thus transmitted are received at the ground stations and the original voltage forms are recovered from the time-modulated signals (Fig. 7). The original data are then separated into various channels, displayed on meters, and recorded by different methods. The purpose of using different methods for recording is to safeguard against the possibility of failure of one method. The principal record is made on a moving strip of photographic paper by means of Hathaway magnetic string oscillograph.

The telemetering equipment for Aerobee is a small (weighing only 20 lbs.), FM-on-FM, 6, channel, 85Mc. unit developed by the A.P.L. [3]. The overall accuracy of the system is claimed to be better than five per cent.

4. MEASUREMENTS AND RESULTS

(a). Ultraviolet solar spectrum

The study of ultraviolet solar spectrum yields data, on the one hand, of the transparency of the terrestrial atmosphere and, on the other, the intensity distribution in the solar spectrum itself. The arrangements of the spectrographs used are briefly described below.

The spectrographs.—The first spectrographs had been designed for the region 3000 A to 2000 A, the portion of the solar spectrum cut off by heavy ozone absorption (see Chap. IV, Sec. 1). The dispersive system was a concave grating (Rowland mounting) with 15,000 lines per inch and 50 cm. radius of curvature.

A difficult problem in the design of the spectrograph is to ensure sufficient illumination on the slit of the spectrograph throughout the course of the flight, notwithstanding the changing aspect of the rocket (with reference to the sun) due to the roll and pitch. The ideal method of overcoming this difficulty is to use sun-follower mirror which automatically so orients itself that the reflected sunlight is always directed on to the slit of the spectrograph. Grating spectrograph with one degree sun-follower system has been successfully designed and constructed by the Applied Physics Laboratory of the Johns Hopkins University.

Two slits, centred on opposite sides of the spectrograph, are used to provide two independent optical paths so that at least one of the slits is always operative. The sun-follower mirror (there are two of them for the two slits) is pivoted about a shaft, the axis of which passes through the plane of the mirror. This axis and that of the rocket are so aligned that solar rays suffer specular reflection when the latter is parallel to the trajectory. The mirror is driven by a motor and its motion is controlled by two photocells placed beside the slit. Light falling on the cells is equal when the slit is fully illuminated. The arrangement is such that the image of the sun is swept across the slit and is held there soon after the mirror becomes illuminated again.

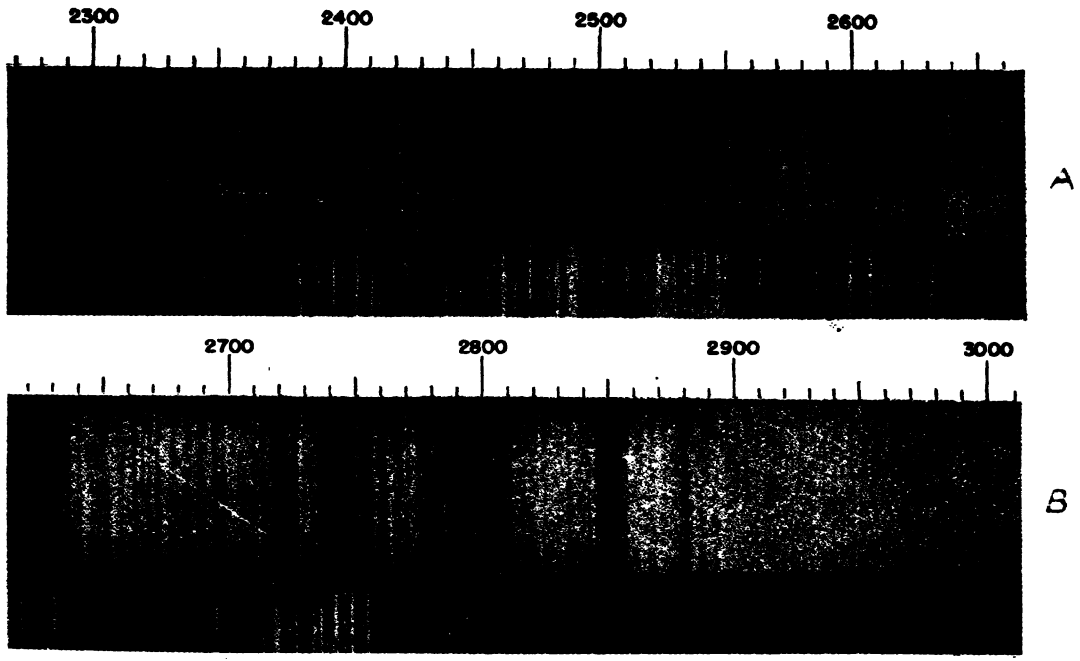
The film in the spectrograph is located on the Rowland circle. There are suitable supply and take up drums. Shutters are provided to exclude light from the film during change, and, with suitable cam, any desired programme of exposures can be obtained. Figs. 8A and 8B, Plate I, are the reproductions of a spectrum photographed at height of 136 km. on July 29, 1947, in a successful V-2 rocket flight.

In an earlier model of the spectrograph (using grating similar to that described above) no sun-follower mirror was used. Instead, the acceptance angle of solar rays was increased by replacing the two slits with two 2 mm. spheres of lithium fluoride (which is highly transparent to the spectral region under consideration). The 'slits' are then the astigmatic images (0·03 mm. in diameter) of the sun formed behind these beads. Two mirrors suitably placed direct the lights coming along the two paths on the concave grating. With such 'bead-slits', though the problem of illumination is partly solved the resolution is necessarily poor as the spectral lines are smeared by the change in the solar angle during an exposure. An automatic motor driven shutter was provided and the film transport mechanism was so arranged at the start of the flight that 0·1, 0·6 and 3 sec. exposures are repeated automatically over and over during the upward flight of the rocket. Fig. 9, Plate II, is a reproduction of a series of photographs taken with this system at different heights on October 10, 1946, in a V-2 rocket flight. A comparison of this series of spectrographs with those reproduced in Figs. 8A and 8B clearly shows the improved resolution in the latter.

Extension of the solar spectrum in the ultraviolet.—Perhaps, the most striking feature of the series of spectra in Fig. 9 is the gradual extension of the spectrum towards the ultraviolet as the rocket rose leaving below more and more ozone. (On account of unfavourable aspect spectrum *G* at 88 km. is weak). Wavelengths in the ultraviolet up to 2100 A were recorded, though, these do not show in the reproduction.

A vertical distribution curve of ozone has been prepared by analysis of these and other similar spectrograms. The general shape of the curve agrees with that obtained by the Umkehr effect [see Chap. IV, Sec. 3(*c*)], but a double peak was observed in the distribution. It is difficult to say, how far the double peak is genuine. No such double peak was observed in the balloon measurements carried out in New Mexico, U.S.A., by Regener [Chap. IV, Sec. 3(*b*)].

The spectrograms in Figs. 8 and 9 show for the first time Fraunhofer lines in the region of solar spectrum cut off by ozone. These are of great interest to the solar physicists. Of these lines, a large number has been identified as due to Fe and Fe^+ (with the help of multiplet table prepared by Mrs. C. E. M. Sitterly [4]). Amongst other conspicuous lines may be noted the resonance absorption line due to $Mg^+(^1P \rightarrow {}^1S)$ at 2852 A and the enormously strong absorption near 2800 A due to the resonance doublet Mg^+ $(^2P \rightarrow {}^2S)$ at λλ 2796–2802. The latter is the analogue of the *H* and *K* lines due to Ca^+ in the visible spectrum. The magnesium doublet is, however, much stronger because of the greater abundance of magnesium. Other



FIGS. 8A and 8B. Plate I—Solar spectrum at a height of 136 km. taken on July 29, 1947. (Iron-arc reference spectrum at the bottom.) A large number of the lines is due to Fe and Fe+. Note the strong absorption line at 2852 A due to Mg and also the very strong absorption at 2800 A due to resonance doublet of Mg+. The prominent line at 2881 A due to Si may also be noted.

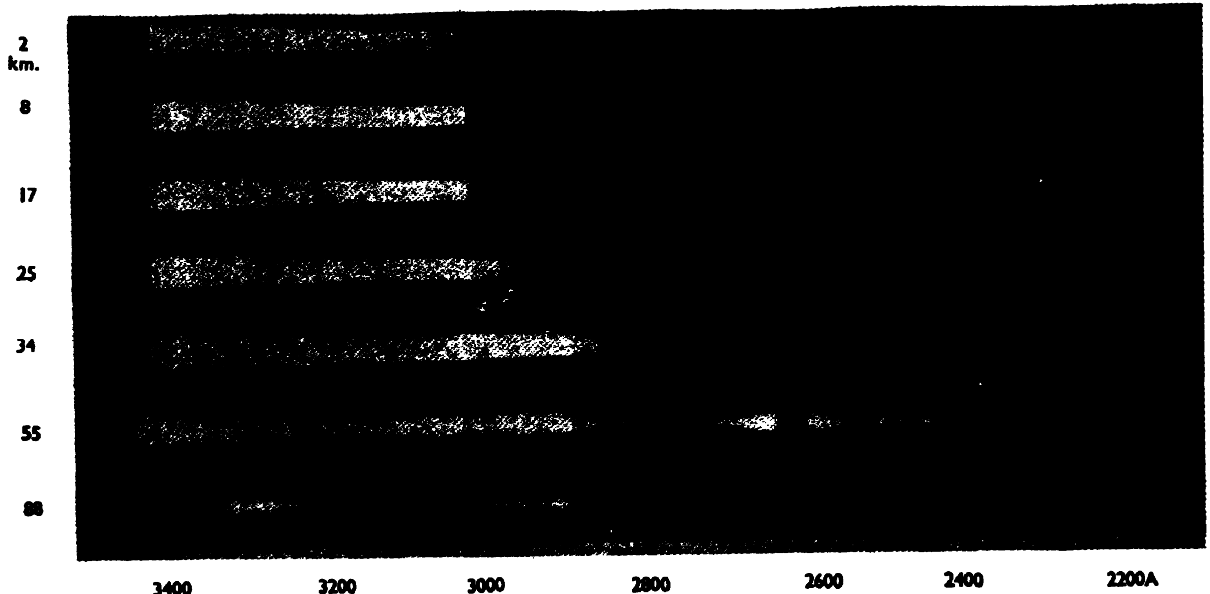


FIG. 9. Plate II.—Series of spectra illustrating the gradual extension of the solar spectrum towards the ultraviolet as the rocket left more and more ozone behind.

The altitudes at which the successive spectrograms were taken, were obtained by knowing accurately the time, after take off, at which each exposure was made. A micro-switch was so arranged in the spectrograph that it opened momentarily each time the film moved. The switch being suitably connected to a telemetering network, a signal was relayed to the telemetering ground station. From these data and the altitude *vs.* time relation, the altitudes were determined.

conspicuous lines identified are at 2510 A and 2881 A (very strong) due to Si and at 2478 A due to C.

The spectral intensity distribution in the wavelength range 3400 A to 2200 A as approximately calculated is shown in Fig. 10. The important

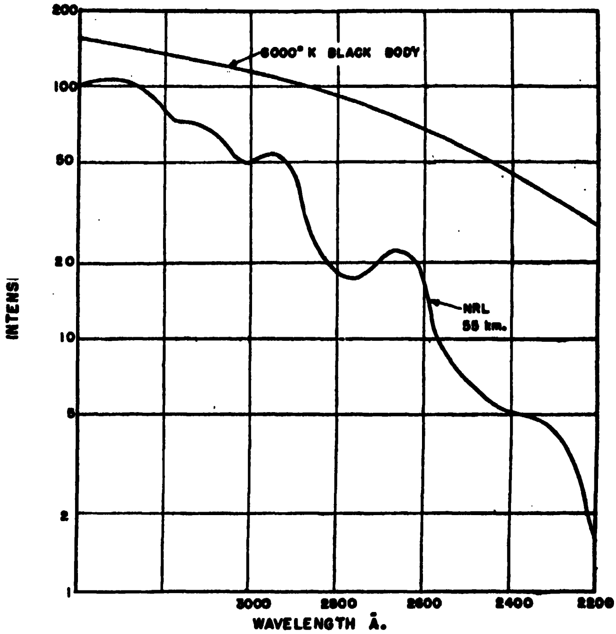


FIG. 10. Intensity distribution in the solar spectrum (lower curve) in the wavelength range 2200 A to 3400 A. Note that the average intensity is significantly below that of the 6000°K black body as usually assumed for the sun.

thing to note in the figure is that the average energy is significantly below that of the ideal 6000°K black body as usually assumed. (Compare Fig. 1, Chap. IV). If this preliminary result is confirmed, then one will have to modify the calculations on ozone distribution and on the ozone heating effect which have been made on the assumption that the intensities in the active wave bands in the solar spectrum are as expected from a 6000°K sun.

(*b*) Pressure measurements

The measuring instruments.—The pressure varies within an enormously wide range—from 665 mm. at the elevation of the site of flight (White Sands Proving Ground, New Mexico) to 10^{-6} mm. at the altitude of 160 km. Different measuring devices are, therefore, used for the different ranges.

For the lower atmosphere in the range 760 mm. to 100 mm. bellows gauges are used. In such gauges the shaft of a 'microtorque' potentiometer is turned by the movements of an aneroid bellows. The potentiometer controls the voltage applied to the telemetering system. Two bellows gauges are used for the above range of pressure. One for 760 mm. to 100 mm. and the other for 100 mm. to 10 mm.

For the range 10 mm. to 0·01 mm. 'hot-wire' or Pirani gauges are used. In this type of gauge the cooling effect of the gas present on an electrically heated wire is measured. The voltage drop across the heated wire element is telemetered. The entire range 10 mm. to 0·01 mm. is covered in three steps.

For the range 10^{-3} mm. to 10^{-5} mm. Philips discharge gauge is used. In this commercial gauge the flow of ions between two electrodes due to the application of 3,000 volt potential between them is increased by a magnetic field. The magnetic field is so inclined to the electric field, that the electrons are compelled to move in helical paths of great length round the field before reaching the anode from the cathode. The chance of ionization by collision is thereby greatly increased and so also the discharge current. The discharge current which passes through a resistor produces a voltage drop and this is telemetered. The entire pressure range (10^{-3} to 10^{-5} mm.) is covered with resistors of different values.

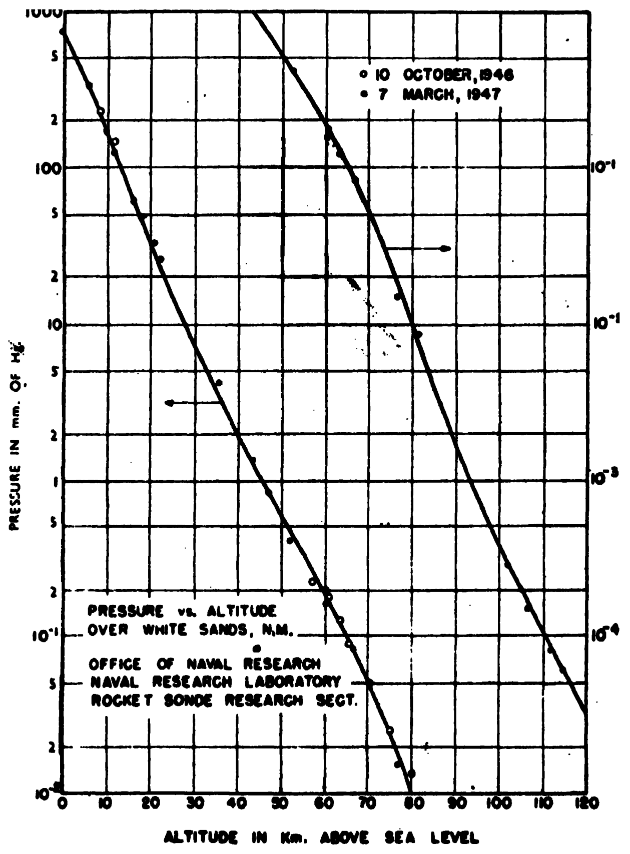


FIG. 11. Pressure *vs.* altitude curve over White Sands Proving Grounds, New Mexico, as deduced from rocket measurements on October 10, 1946, and March 7, 1947.

The ambient pressure, as also the so-called ram pressure (pressure of the air which has been compressed at the front, e.g. an opening in the nose, due to the high velocity of the rocket) have been measured. The former, up to about 80 mm., was measured by Pirani gauges placed outside the rocket (V-2) just forward of the tail section. The pressure here was very nearly equal to the ambient pressure as checked by wind-tunnel test. The latter was measured at the nose of the war-head by Philips gauge. The ambient pressure could also be deduced from the ram pressure by the theory of Taylor and Maccoll. This gave pressure values up to 120 km.

The results of pressure measurement as made in two successful V-2 rocket flights on October 10, 1946, and March 7, 1947, are depicted in Fig. 11.

For the portion below 25 km., the pressure was obtained from balloon data taken within one hour of the rocket flight. These are included to complete the curve. The pressure values are within the limits of variation (diurnal and seasonal) as predicted in the Tables published by the National Advisory Committee for Aeronautics (N.A.C.A.).

(*c*) Temperature measurements (indirect)

On account of the velocity of the rocket being considerably greater than the average velocity of air-molecules it is not possible to make direct measurement of air temperature. (At the instant the fuel is burnt out, the velocity is three times the average molecular velocity for the V-2 rocket.) At such velocities the temperature of the air mass to be measured is profoundly modified. Further, with the decrease of density with height the heat content of the atmosphere also decreases. Hence, for the great heights (above 60 km.,) the temperature of the measuring thermometric elements, even with careful shielding, is controlled by radiation and not by heat transfer by air conduction. For these reasons the temperature is determined indirectly from the data of pressure measurements. Two methods of utilizing the pressure data have been successfully employed.

The first is from the slope of the log ρ-h curve. We have from Eq. (7), Chap. I, Sec. 3(*b*),

$$\log\left(\frac{\rho}{\rho_0}\right) = -\frac{h}{H}.$$

The negative of the slope of the log ρ-h curve is thus proportional to the scale height $H = kT/mg$. Now, as discussed in Chap. I, Sec. 4, the composition of the atmosphere may be assumed to remain unaltered up to at least 80 km. level. (This conclusion is corroborated by the results of analysis of air samples collected in rocket flights at the height of 70 km. *Vide infra.*) Hence, if m is taken to be the mean molecular mass of air, the temperature at any height can be calculated. A close inspection of the curve in Fig. 11 will reveal the positions where there are changes in the slope. The heights at which such changes are noticed may be compared with the heights where the temperature gradient changes.

The second method is to compare the ram pressure with the ambient pressure. The ratio of the two pressures depends in a known manner on the ratio of the velocity of the rocket to that of sound, the so-called Mach number (M). For large values of M this ratio is equal to $1 \cdot 29\ M^2$. Hence, from the value of M (as calculated from the ratio of the two pressures) and from the velocity of the rocket (which is tracked throughout the flight) the veloity of sound at any height is determined. This velocity being proportional to $\sqrt{T/mg}$, the temperature is easily calculated. In Fig. 12 the

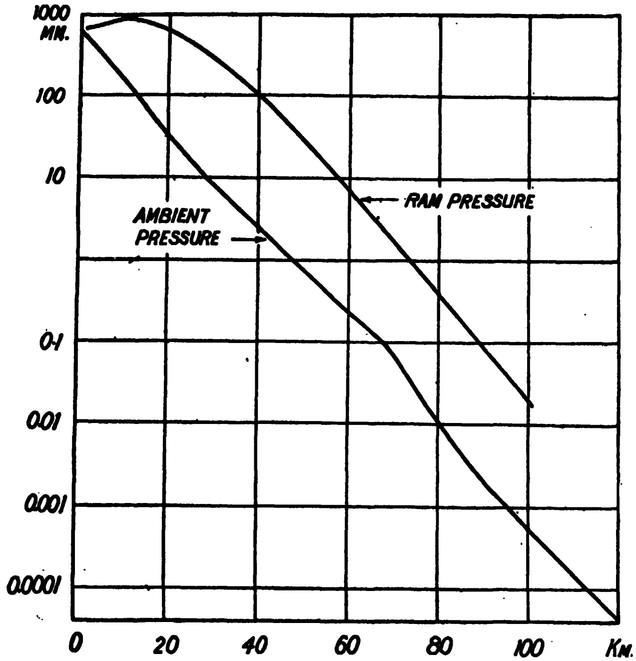


FIG. 12. Illustrating the variations of ram pressure and ambient pressure as recorded in the V-2 rocket flight on March 7, 1947. (Only the smoothed-out curves have been drawn omitting the recorded points.) Records in the tropospheric region are from those obtained in a balloon flight made at about the same time.

graphs of the variations of the ambient pressure and ram pressure are given. The data for ambient pressure, up to about 15 km. are from balloon recordings.

The temperature curve as obtained by the above method [5] is shown in Fig. 1, p. 581, Chap. XIII. It should be mentioned that the probable errors of measurement are rather large being ±20°C. from 50 to 60 km., ±15°C. at 65 to 70 km. and ±20°C. at 72 to 75 km. For regions above 100 km., the error is ±40°C.

The same pressure data have also been examined by Nazarek [6]. The most important result obtained by this author is that there appears to be an irregularity near 100 km., (in the transition layer $O_2 \rightarrow O+O$) where the temperature decreases slightly.

It may be noted that though it had not been possible to make direct measurements of ambient temperature, the skin temperature of the rocket near the nose has been measured with the help of platinum resistance thermometer. Thus, the rise in temperature in V-2 rockets are 120°$\pm$ 5°C. on the 0·1 inch thick aluminium forward section, and 85°$\pm$5°C. on the 0·1 thick steel immediately behind the aluminium section.

(*d*) Ionosphere experiments

These experiments were designed to determine the change in the refractive index of the atmosphere at ionospheric heights due to the influence of electrons and ions present there. The method consists in observing the difference in the velocity of propagation of two waves, one on 4·5 Mc./s. (which was close to the critical frequency of the *E*-layer at the time of the flight) and the other near 25 Mc./s. which is so far removed from this frequency that its velocity is not affected by the ionization. Waves on these two frequencies are sent out from the rocket, the higher frequency being obtained by a six-fold frequency multiplication of the lower.

At the ground station the lower frequency after being received is multiplied six times and mixed with the higher frequency in a beat frequency detector. In observations taken in course of a V-2 rocket flight on March 7, 1947, it was found that as long as the rocket was in the region below the *E*-layer (in which there is no large ionization) there were no beats. But as the rocket entered the *E*-layer beats, whose frequency increased progressively, were obtained up to an altitude of 110 km. The beats disappeared at this height and were again obtained for a considerable time on the downward flight.

Analysis of such data, so far as have been made, have yielded estimates of the effective, rather than the actual, electron number densities. Thus, the preliminary analysis of data obtained in a flight on January 22, 1948, shows that for the region below 85 km. the effective electron number density was less than 10^4/cm.^{3} rising to about $2 \cdot 3 \times 10^5$/cm.^{3} at 100 km. The density was appreciably less than 10^5/cm.^{3} at 130 km. and appeared to have again increased at 155 km. [2].

(*e*) Analysing composition of the Upper Atmosphere

The chemical composition of air at 70 km. height has been determined and has been found to be practically the same as that in the troposphere.

For such determinations samples of air have to be collected when the rocket is in free flight (i.e. after the fuel is burnt out). A method of doing this, as has been successfully developed by the Department of Aerodynamics of the University of Michigan, U.S.A., is as follows. The air is collected in steel bottles which had been evacuated prior to their being fitted in the rocket. The bottles are connected to a 'flushing-chamber' through a thin-walled tinned copper tube and a break-seal. Ambient air, from an inlet port in the side of the rocket, streams through the flushing-chamber and is discharged out of a 'reverse-scoop' nearer the tail. The break-seal

is automatically opened at a pre-determined instant of flight when air from the flushing-chamber enters the evacuated steel bottles. After a few seconds the bottles are automatically sealed off by a device which heats and simultaneously squeezes the copper tube. After recovery the gases from the bottles are pumped into lime-glass bulbs for distribution to the organizations which are to carry out the analysis.

The gases of a few such bottles (collected in V-2 rocket flights) were analyzed by Chackett, Paneth and Wilson [7] with the result as described above. In the actual analysis the ratios of helium and neon to nitrogen (plus argon), as also the ratio of argon to nitrogen (plus argon) were determined. No significant departure from the proportion in which these gases exist in the ground-level air was detected. This shows that there is no diffusive separation of gases at least up to this height. (See Chap. I, Sec. 4.) It is interesting to note that the bottles did not contain any oxygen. This is because it was removed by combination with copper of the copper tube (connecting the bottles with the flushing chamber) when it was heated for the seal off. In one of the samples analysis was made for hydrogen. But it was not found in detectable quantity, i.e. it was less than 0·1 per cent. (See Chap. X, Sec. 8.)

(*f*) Aerial Photography

One of the main objects of taking the photographs is to secure information regarding the various motions executed by the rocket in course of its flight. Amongst other purposes may also be mentioned information about corona discharge, if any, from the radiating elements. Fig. 13 is the reproduction of one of a series of photographs taken from V-2 rocket on March 7, 1947, fired from the White Sands Proving Grounds, New Mexico. The photograph was taken from an altitude of 162 km. with K-25 aircraft camera. There were two such cameras mounted symmetrically on opposite sides of the central section of the rocket. The field of view of the cameras (provided with right-angled prisms) was directed towards the tail of the rocket making an angle of 20° with its axis. The photograph shows the aspect of the country towards the south-west from the rocket. The Gulf of California and the peninsula of Lower California are clearly seen. The curved horizon is 1,500 km. towards the west and the photograph covers more than $5{\cdot}18 \times 10^5$ sq. km. of United States and Mexico.

(*g*) Magnetic field at high altitude—evidence of upper atmospheric current sheet

The terrestrial magnetic field at high altitude (100 km. above sea-level) has been measured with a view to obtain evidence of the world-wide electric current sheets which cause quiet-day terrestrial magnetic variations. (See Chap. VII, Sec. 5.)

A total-field magnetometer, whose 'readings' are independent of its orientation has been specially designed and built for the purpose by the Naval Ordnance Laboratory and the Bell Telephone Laboratories. The

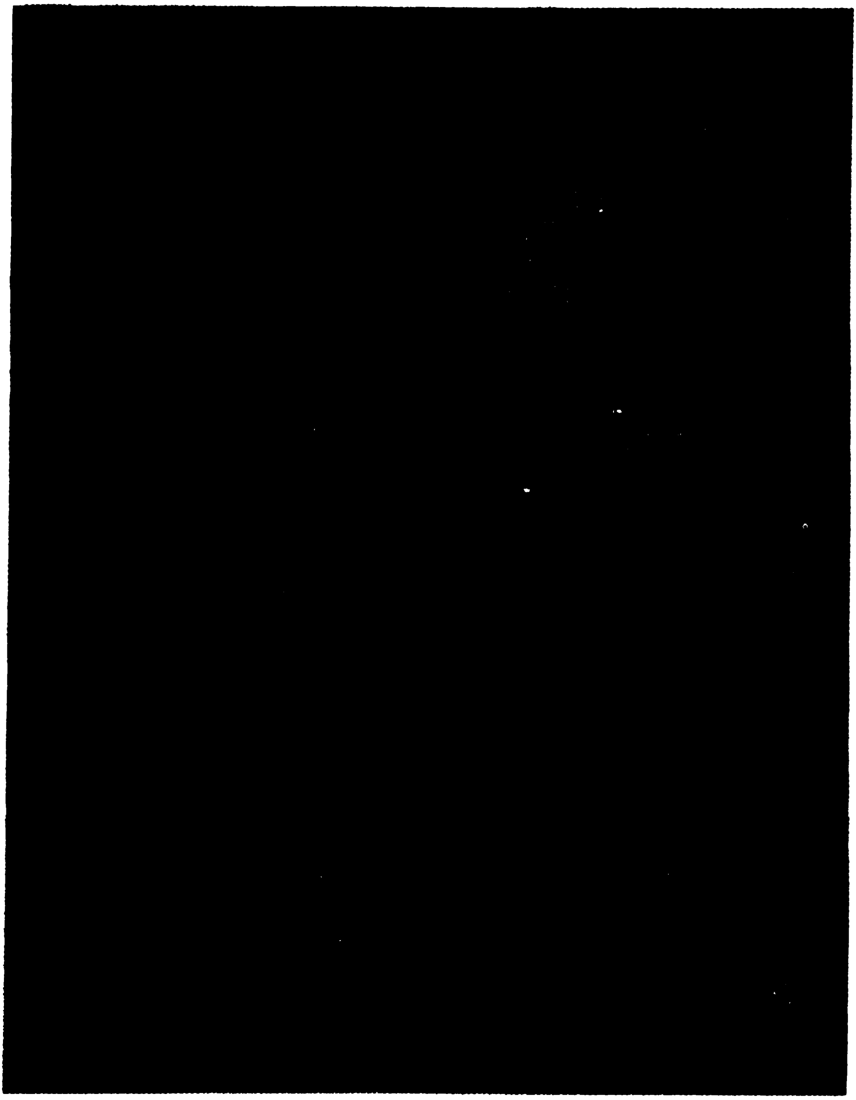


FIG. 13. Photograph of the earth surface taken from an altitude of 162 km. over the White Sands Proving Grounds, New Mexico, U.S.A., on March 7, 1947. The photograph shows the aspect of the country towards the south-west. The gulf of California and the peninsula of Lower California are clearly seen. The curved horizon is 1,500 km. towards the west and the photograph covers more than $5{\cdot}18 \times 10^5$ sq. km. of United States and Mexico. (Official United States Naval Photograph.)

magnetometer consists of three mutually perpendicular coils wound on permalloy. These are driven by an oscillator and the output voltages of the three elements are squared and added. The resultant signal is thus proportional to the square of the magnetic field acting on the coils. The

magnetometer, together with the power supply (storage batteries and dynamotors), photo-orienter (for indicating the aspect of the rocket with respect to the sun), cut-off receiver, pressure gauges and the telemetering system are installed in the forward section of an Aerobee rocket.

A first attempt at obtaining evidence of the upper atmospheric current sheet was made at White Sands Proving Grounds, New Mexico, on April 13, 1948, at 1441 hours M.S.T. The height reached was 113 km. above sea-level (112 km. above ground level), and, though the variation of total field intensity with height was found to agree closely with that calculated from the inverse-cube-law, no evidence of the current layer was obtained. The reason for this failure was ascribed to the unfavourable site of the rocket flight [8]. The expected 'discontinuity' in the total field due to current sheet at this site (41° North magnetic latitude) is only 10 to 20 gammas, while the decrease in the total magnetic field intensity due to the inverse-cube-law is 28 milligauss. (1 gamma = 10^{-5} gauss.)

However, later attempts made at more favourable location near the magnetic equator have proved successful [9]. At such latitudes the expected discontinuity is 200 to 220 gammas, and decrease in field intensity due to the inverse-cube-law is only 13 milligauss.

The Aerobee sounding rocket (A-11), with the magnetometer mounted as described above, was fired on March 22, 1949, at 1120 hours (90th meridian time), when the diurnal magnetic variation at the surface is about the maximum, from the seaplane tender USS Norton Sound, off the west coast of Peru (geomagnetic latitude −1°, geomagnetic longitude 341°). A decrease in the field intensity of about 400 gammas between 93 km. and 105 km. was obtained, in addition to the inverse-cube-law decrease. This decrease corresponds fairly well with that calculated from the diurnal variations recorded at the ground. More accurate agreement could not be obtained as a magnetic storm was in progress at that time.

It is interesting to note that the records of a similar Aerobee flight made a few days earlier (March 17, 1949) from the same place but at 1720 hours, did not show any discontinuity. This is as expected. The diurnal variation of the terrestrial magnetic field (and, hence the overhead current intensity) is small at this hour. (See Fig. 12, Chapter VII.)

The results of these flights provide direct experimental evidence of the existence of electric current systems in the E-region of the ionosphere and thus support the dynamo theory of quiet day terrestrial magnetic variations (Chapter VII, Sec. 5).

(h) Other investigations

Among other investigations that have been carried out with rocket-borne instruments mention may be made of measurements on cosmic rays, on day air-glow and scattered light and on the solar constant.

(i) *Cosmic rays.*—Rockets have carried cosmic ray measuring apparatus to heights up to about 165 km. However, in spite of such great heights attained, rocket exploration has not been able to add as much

information on cosmic rays as desired. This is because the rocket stays above the 'top' of the atmosphere for a very short time (only a few minutes) and also because scattering and shower formation on the walls of the rocket interferes with the observation. As such, sounding balloons which go only up to 30–35 km., but may stay up to 8 hrs., are still the most useful agency employed for cosmic ray study. The important result that has been obtained by rocket observation is that the counting rate with Geiger counters after attaining a maximum at about 20 km. (as observed in balloon experiments) decreased and remained practically constant from a height of 55 km. This shows that above 55 km. there are only primaries and that the 'top' of the atmosphere had been reached.

The primary particles at the 'top' of the atmosphere consist mostly of protons arriving isotropically from all directions. Balloon experiments have also shown evidence of multiply-charged nuclei, stripped of outer electrons, above the 'top' of the atmosphere. It has also been found that the cosmic ray radiation at the 'top' depends to some extent on the latitude of the place of observation, the relatively low energy particles being deflected towards the terrestrial magnetic poles by the combined actions of the solar and terrestrial magnetic fields [10].

It is interesting to note that suggestion has been made that variation of cosmic ray intensity with altitude may be utilized to determine the rocket trajectory [11].

(ii) *Day air-glow.*—Preliminary measurements have been made on day air-glow, at 136 km. height with monochromatic photoelectric photometer. The results show that the value of the day air-glow is much larger than that predicted by the existing theories. Further experiments on the subject are in progress.

(iii) *Solar constant.*—Intensity of the solar radiation has been measured at an altitude of 65 km. by the University of Rhode Island, under contract with the U.S. Air Force, using a 'semi-conductor' type of bolometer (thermistor bolometer) as detector of the radiation. The preliminary results obtained show that the values cluster round 2 gram calories per square centimeter per minute (see Appendix, Sec. 1). It is planned to improve the apparatus in future measurements.

(iv) *Wind and temperature measurements.*—Specially generated smoke clouds have been used by the U.S.A. Signal Corps to determine winds in the stratospheric regions.

The temperature in the region 50–60 km. has also been measured by the Signal Corps by measuring the velocity of sound produced by firing of explosive charges when the rocket is on the ascent [2].

(v) *Suggested investigations.*—Other experiments with rocket-borne equipment that suggest themselves are: direct spectroscopic observation of the variation of atmospheric absorption with height and location of the emitting layers in night air-glow. O'Day has suggested that high-frequency high-tension coronal discharge may be produced in the surrounding atmosphere from the nose of the rocket and that the spectrum of the

discharge glow could be recorded by a fast spectrogram suitably located within the rocket.

It can be confidently predicted that with improved types of rockets and with improved instrumentation and recording technique many of the physical characteristics of the upper atmosphere, our knowledge about which is poor and which is gained only by indirect methods, will be directly measured and so yield sure and complete data.

CHAPTER XIII

CONCLUDING REMARKS

1. SUMMARY

In concluding the treatise it would be useful to give a summary of the contemporary state of our knowledge of the physical state of the upper atmospheric regions.

Composition.—The atmospheric composition in the troposphere, according to the best available data is as shown below [1]:

(per cent by volume)

Nitrogen	78·08	99·9 (Nitrogen, Oxygen, Argon, Carbon dioxide)
Oxygen	20·95	
Argon	0·93	
Carbon dioxide	0·03	
Neon	$1{\cdot}8\times10^{-3}$	
Helium	5×10^{-4}	
Krypton	1×10^{-4}	
Xenon	1×10^{-5}	
Ozone	Variable : $>1\times10^{-6}$	
Radon (average near ground)	6×10^{-18}	
Hydrogen	Doubtful : $<1\times10^{-3}$	

This composition, at least insofar as the principal gases oxygen and nitrogen, and the minor gases argon and helium are concerned, is maintained at least up to the height 70 km. [2].

The atmosphere in the region 10–50 km. is comparatively rich in ozone, the maximum concentration being in the region 25–30 km. The ratio of ozone to air by volume is, however, maximum at a higher level at about 35 km. (Chapter IV, Fig. 12). The total ozone content is quite small being only equivalent to 0·1 to 0·2 cm. when reduced to the standard temperature and pressure, but, it suffices to cut off entirely by absorption the solar spectrum between 2900 A and 2100 A.

There are reasons to believe that in the middle atmosphere round 67 km. there is a 'layer' of molecular hydrogen, with a concentration of about 10^{10} molecules/cm^3. (Chapter X, Sec. 8).

Spectroscopic evidences have been obtained of the presence of the OH radical and of Na atoms in the upper atmosphere [Chapter X, Sec. 4(*c*)]. The former is concentrated more or less at about the height 70 km. (as estimated from measurement of the height of the 'layer' emitting OH bands). There is a 'layer' of neutral sodium atoms in the region 35–75 km. Sodium is also present in the higher regions, but above 100 km. it is almost

entirely ionized. The total quantity of sodium is, however, extremely small, namely, only about 10^{-12} of the atmosphere.

Above 90 km. the atmospheric composition changes from one of N_2 and O_2 to one of N_2 and O due to the dissociation of $O_2 \rightarrow O+O$ by absorption of ultraviolet radiation ($\lambda < 1760$ A). Above 130 km. the concentration of O_2-molecules falls to a negligible value as compared to that of O-atoms (Chapter V, Fig. 1).

There is also spectroscopic evidence of the dissociation of $N_2 \rightarrow N+N$ in the upper atmosphere, at least, in the regions of auroral displays. However, atomic nitrogen may not be as abundantly present and as widely distributed as atomic oxygen (Chapter V, Sec. 3).

The Ionospheric regions.—The atmospheric constituents from above 60 km. are more or less ionized (Chapter VI, Fig. 1).

The regions of ionization—known collectively as the *ionosphere*—can 'reflect' radiowaves and thus play a fundamental rôle in the transmission of radio signals round the world. There are four main regions of ionization, designated D, E, F_1 and F_2. Region E has a well defined ionization maximum and is situated at about 100 km. height. Region D is present only during daytime. It has no such maximum and is just a bank of ionization between 60 km. and 100 km. The maxima of regions F_1 and F_2 lie approximately at heights of 200 and 275 km. respectively. During night the two regions F_1 and F_2 merge together to form a single region of ionization called F.

The upper atmospheric ionization is strongly under solar control, the degree of ionization varying with the hour of the day, with the season of the year and with the epoch of the solar cycle. The average maximum electron concentrations in regions E and F, during the epoch between the maximum and minimum of solar activity, are of the orders 10^5 and 10^6 electrons/cm.^{3} respectively. The electron concentration in the D-region at 60 km. height is of the order 2×10^2/cm^3. Higher up, where it merges with the tail of the E-region, it is of the order 10^4/cm^3. The D-region is mainly an absorbing region, but is also helpful in the propagation of long- and very-long-waves.

The ionization is also strongly affected by bombardment of solar corpuscles as produce magnetic storms and auroras. During magnetic storms and auroral displays the ionospheric condition, particularly in the high magnetic latitudes, is greatly disturbed.

The production of the different ionospheric regions is ascribed to the ionizations of the different atmospheric constituents. Thus, Region E is believed to be produced in the region of transition $O_2 \rightarrow O+O$ by preionization of O_2 (Chapter VI, Fig. 66). Region F_1 is produced by the ionization of O and Region D by ionization of O_2 at their first ionization potentials. For F_2 no separate ionization process is assumed. It is believed to be produced by a sort of bifurcation of Region F_1. [See Chapter VI, Sec. 10(d).]

The ionized regions being traversed by the terrestrial magnetic field exhibit the phenomenon of magnetic double refraction.

Temperature distribution.—The temperature distribution curve as shown in the left in Fig. 1 was obtained from balloon and rocket data (see Chapter XII, Sec. 4c). The curve on the right is the standard distribution as

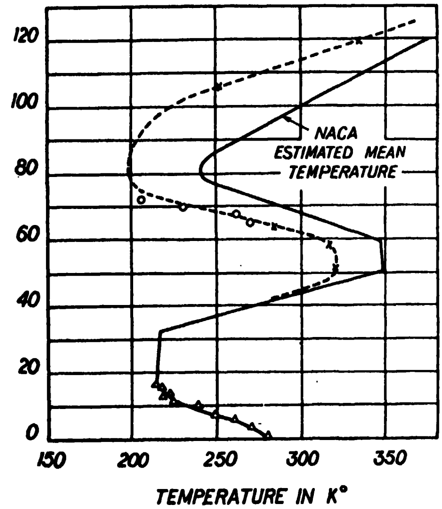


FIG. 1. Atmospheric temperature distribution with height. *Dashed-line curve on the left*—Temperature distribution with height as obtained from rocket data (March 7, 1947). *Crosses*—data calculated from slope of pressure curve. *Circles*—data calculated from ram pressure. The data for the tropospheric regions (*triangles*) were obtained from a balloon flight. *Continuous-line curve on the right*—Standard temperature distribution as adopted by the National Advisory Committee for Aeronautics, U.S.A.

adopted by the National Advisory Committee for Aeronautics (N.A.C.A.) of the U.S.A. (The height-temperature distribution as depicted in these curves was first obtained by Martyn and Pulley from various considerations [3].) The drop in the temperature in the section 0–18 km. and its subsequent constancy in section 18–32 km. at the base of the stratosphere are, as has been known for a long time, the effects of convective and radiative equilibriums respectively in these regions. [See Chapter I, Sec. 3(*c*).] The rise in the middle atmosphere in the section 32–55 km. is due to absorption of solar radiation by ozone in the region of the Hartley bands (Chapter IV). Observations on the abnormal propagation of sound waves and critical study of the heights of appearance and disappearance of meteor flashes confirm this (Chapters IIIA and IIIB). The drop at 80 km. is to be ascribed to the fact that there is no absorption of solar radiation in this region. The low temperature is inferred from the occurrence of noctilucent clouds within a narrow range in this region. [See Chapter VI, Sec. 14(*b*), Fig. 83, Plate IV.] These clouds consisting of ice crystals are often seen in northern countries by being illuminated by the oblique rays of the sun from below horizon. The low temperature is also confirmed by radio measurements of the scale height H $(= kT/mg)$. The rise in temperature from above 85 km. is ascribed to absorption of solar radiation by O_2. The temperature attained at 300 km. level is variously estimated to lie between 500° and 3000°K. The higher values are probably attained only during solar flares [Chapter I, Sec. 5(*b*)]. It is probable that above 300-400 km. the temperature falls gradually to merge with that of interstellar space.

Atmospheric density-variation with height.—Results of calculations of air density made on the assumption of a temperature distribution of the

TABLE I

Distribution of molecular number density with height

Height (km.)	Temp. (°K)	Total pressure (mm.)	Density (number of particles per c.c.)			Remarks.
			N_2	O_2	O	
0	297	760·0	$2{\cdot}01\times10^{19}$	$5{\cdot}00\times10^{18}$	..	N_2 and O_2 thoroughly mixed. Temperature distribution approximates to that given in Fig. 1.
10	232	210·3	$7{\cdot}16\times10^{18}$	$1{\cdot}79\times10^{18}$	..	
20	220	41·77	$1{\cdot}48\times10^{18}$	$3{\cdot}70\times10^{17}$	..	
30	220	9·35	$3{\cdot}34\times10^{17}$	$8{\cdot}34\times10^{16}$	..	
40	240	2·05	$6{\cdot}68\times10^{16}$	$1{\cdot}67\times10^{16}$	..	
50	330	0·78	$1{\cdot}85\times10^{16}$	$4{\cdot}62\times10^{15}$	..	
60	420	0·23	$4{\cdot}28\times10^{15}$	$1{\cdot}07\times10^{15}$	..	
70	450	0·11	$1{\cdot}86\times10^{15}$	$4{\cdot}68\times10^{14}$	..	
80	160	$3{\cdot}14\times10^{-2}$	$1{\cdot}54\times10^{15}$	$3{\cdot}84\times10^{14}$	..	
90	200	$4{\cdot}71\times10^{-3}$	$1{\cdot}95\times10^{14}$	$4{\cdot}78\times10^{13}$	$2{\cdot}71\times10^{6}$	Region of dissociation of O_2 to O. Temperature increase is taken to be 4°K per km. (See Chapter V.)
100	240	$1{\cdot}00\times10^{-3}$	$2{\cdot}82\times10^{13}$	$7{\cdot}56\times10^{12}$	$1{\cdot}09\times10^{12}$	
110	280	$2{\cdot}70\times10^{-4}$	$8{\cdot}58\times10^{12}$	$8{\cdot}25\times10^{10}$	$3{\cdot}59\times10^{12}$	
120	320	$8{\cdot}68\times10^{-5}$	$2{\cdot}54\times10^{12}$	$5{\cdot}29\times10^{8}$	$1{\cdot}06\times10^{12}$	
130	360	$3{\cdot}19\times10^{-5}$	$8{\cdot}60\times10^{11}$	$1{\cdot}72\times10^{7}$	$3{\cdot}45\times10^{11}$	
150	440	$8{\cdot}71\times10^{-6}$	$1{\cdot}47\times10^{11}$	..	$1{\cdot}08\times10^{11}$	Oxygen in atomic state only. Diffusive separation of O and N_2. Temperature increase same as above.
175	540	$2{\cdot}27\times10^{-6}$	$2{\cdot}78\times10^{10}$	..	$3{\cdot}28\times10^{10}$	
200	640	$7{\cdot}51\times10^{-7}$	$4{\cdot}31\times10^{9}$	..	$1{\cdot}75\times10^{10}$	
250	840	$1{\cdot}28\times10^{-7}$	$3{\cdot}28\times10^{8}$	..	$2{\cdot}55\times10^{9}$	
300	1040	$3{\cdot}83\times10^{-8}$	$4{\cdot}71\times10^{7}$	..	$7{\cdot}15\times10^{8}$	
400	1440	$7{\cdot}23\times10^{-9}$	$1{\cdot}93\times10^{6}$	..	$1{\cdot}08\times10^{8}$	
500	1840	$2{\cdot}25\times10^{-9}$	..	..	$2{\cdot}63\times10^{7}$	
600	2240	$8{\cdot}86\times10^{-10}$	..	..	$8{\cdot}49\times10^{6}$	
700	2640	$6{\cdot}12\times10^{-10}$	..	..	$3{\cdot}41\times10^{6}$	
800	3040	$2{\cdot}13\times10^{-10}$	..	..	$1{\cdot}38\times10^{6}$	
900	3440	$1{\cdot}26\times10^{-10}$	..	..	$7{\cdot}33\times10^{5}$	

type shown in Fig. 1 is given in Table I. (The assumptions regarding the degree of mixing at different heights are shown in the last column of the Table.) The figures for heights up to 100 km. may be taken as fairly representative of the average condition. In support of this it may be noted that the calculated pressure at 90 km. is quite close to that deduced from radio observations [Chapter VI, Sec. 15(iv)]. The figures for regions above 100 km. level are, however, somewhat speculative on account of the uncertainties in the nature of the assumptions. With different assumptions the distribution of particle concentration with height will be different. As an example we give in Table II, the distribution for the 'model atmosphere' after Gerson [3*a*]. The assumptions made in the computation of the Table are as follows:

Temperature distribution with altitude is as shown in Fig. 2 and holds at lat. 45°N. for January and August conditions respectively.

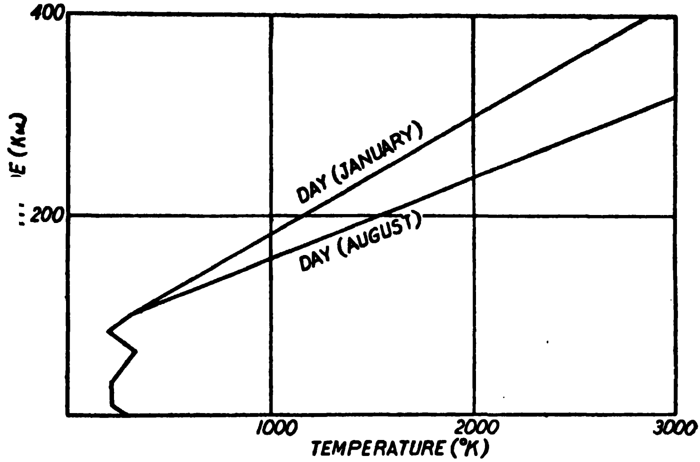


FIG. 2. Altitude-temperature distribution in 'model atmosphere' for conditions prevailing in 45°N., January and August respectively. (After Gerson.)

Oxygen dissociation starts at 94 km. level and is 100% at 100 km. level the dissociation increasing linearly.

Complete mixing of the atmospheric constituents at all heights.

It will be noticed that the distribution of the particle concentrations with height up to 100 km. is practically the same in the two Tables. This is because of the nearly identical altitude-temperature distribution assumed up to this height. But, above 100 km. level the figures in the two Tables differ even in the order of magnitude.

TABLE II

Model Atmosphere

(For the assumptions made in the computations see text)

Height (km.)	Temperature (°K)	Pressure (mm.)	Number density $= P/kT$	Classical mean free path (cm.)
0	288	760	$2{\cdot}549 \times 10^{19}$	$8{\cdot}625 \times 10^{-6}$
10·8	218	176·2	$7{\cdot}805 \times 10^{18}$	$2{\cdot}816 \times 10^{-5}$
32	218	6·46	$2{\cdot}860 \times 10^{17}$	$7{\cdot}685 \times 10^{-4}$
62	330	$1{\cdot}53 \times 10^{-1}$	$4{\cdot}480 \times 10^{15}$	$4{\cdot}906 \times 10^{-2}$
84	200	$9{\cdot}03 \times 10^{-3}$	$4{\cdot}361 \times 10^{14}$	$5{\cdot}041 \times 10^{-1}$
94	262·5	$2{\cdot}13 \times 10^{-3}$	$7{\cdot}821 \times 10^{13}$	2·811
100	300	$1{\cdot}12 \times 10^{-3}$	$3{\cdot}593 \times 10^{13}$	6·118
300 August	2700	$8{\cdot}57 \times 10^{-6}$	$3{\cdot}064 \times 10^{10}$	$7{\cdot}174 \times 10^{3}$
January	2000	$2{\cdot}99 \times 10^{-6}$	$1{\cdot}443 \times 10^{10}$	$1{\cdot}542 \times 10^{4}$
400 August	3900	$4{\cdot}27 \times 10^{-6}$	$1{\cdot}057 \times 10^{10}$	$2{\cdot}079 \times 10^{4}$
January	2850	$1{\cdot}16 \times 10^{-6}$	$3{\cdot}929 \times 10^{9}$	$5{\cdot}594 \times 10^{4}$

Tides and winds.—The atmosphere, like the oceans, is subject to large scale oscillations (tidal) due to the gravitational pulls of the sun and the moon (Chapter II). There are, however, two significant differences. Firstly, owing to the temperature drop at 60–80 km. level the atmosphere, as a

whole, has more than one mode of oscillations. Besides the mode with the period of 10½ hours (as excited by catastrophic explosions) there is, amongst others, one with period of nearly 12 hours, that is, almost exactly equal to half the solar day. As a result of this the solar tidal oscillation (in contrast to those in oceans), is greatly enhanced by resonance and exceeds the lunar oscillation by sixteen times. Secondly, the heating effect of the sun excites tide-like oscillations and contributes about 50% to the total of the solar oscillation. The solar oscillations manifest themselves by rhythmic daily rise and fall (amplitude of the order 1 mm.) of the barometric pressure. The lunar oscillations are not so easily observable, but are detected only after laborious statistical analysis of many years' data.

The upper atmospheric regions are also subject to winds. The primary cause of such wind circulation (apart from tidal forces which cause wind-like motions) is absorption and radiation of energy derived from the sun by agents like ozone in the middle atmosphere and oxygen (and, probably also, nitrogen) in the ionospheric regions. The absorption heats the atmospheric gases, and, the temperature attained is determined partly by radiation and partly by conduction and transportation of heat into and out of the heated gas. Of these, effect of conduction may be negligible. But transport of heat, due to convection, turbulence and general circulation, may be quite important and be the decisive factor in determining the temperature.

Now, in Fig. 2 the temperature profile of the atmosphere has been drawn, showing the vertical gradients, up to the ionospheric regions. A quite large seasonal variation of temperature has been shown in the ionospheric regions. Such seasonal variations also exist in the middle atmospheric regions as evidenced by abnormal wind propagation data and by observations on meteor trails. There are thus meridional temperature gradients at different levels, for example, when it is summer in northern hemisphere and winter in the southern and vice-versa. These temperature gradients—vertical and horizontal—are the immediate causes of the establishment of general wind circulation; for, there must be transport of heat from regions of higher temperature to those of lower temperature. (It is to be noted that the effect—wind circulation—also influences the cause—temperature gradient. But, this may be neglected in a first order approximation.) The circulation, as in the lower atmosphere, is controlled by the rotation of the earth (deflecting the north-south flow into an east-west flow), by the conservation of angular momentum (the momentum being transported poleward or equatorward according as there is a northerly or southerly meridional exchange of air mass) and, in the highest ionospheric regions, by internal friction arising out of viscocity (see Fig. 3, Chap. I) producing large wind components across the isobars.

The methods direct and indirect by which wind data in the upper atmospheric regions have been collected, have been discussed in the preceding chapters in appropriate places. These are: Balloons, smoke shells and rockets; abnormal sound wave propagation phenomena; obser-

vations on meteor trails and noctilucent clouds; radio-pulse explorations; studies of barometric oscillations and geomagnetic variations.

Surveys of the contemporary state of our knowledge—both theoretical and experimental—of the general circulation of wind in the upper atmospheric regions have been made by Kellog and Schilling [4] and also by Wan-Cheng Chiu [4*a*]. The former authors have also prepared a schematic representation of the meridional distribution of the wind system up to 120 km. Upper atmospheric wind circulation has also been considered by Yerg [4*b*]. According to this author the viscous action is maximum at 200 km. altitude. Taking this and the observed temperature lapse rates into consideration a simple meridional circulation is developed with a northward flow below and a southward flow above 200 km. In Fig. 3 we reproduce tentative curves of variations of wind-velocity (easterly or

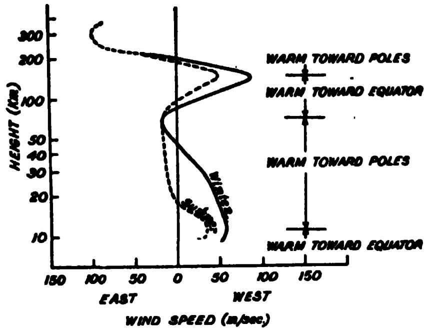


FIG. 3. Estimated distribution of wind (east and west components) with height in middle latitudes for summer and winter seasons. The height is given in logarithmic scale to enable one to infer directly the approximate mass flow. On the right, the meridional temperature gradient for quasi-geostrophic winds is shown. (After Sheppard.)

westerly) with height, for middle latitudes in northern hemisphere, prepared somewhat earlier by Sheppard [4*c*]. The observations on which the curves are based are as follows:

In the troposphere a fairly well defined wind with a maximum of 30 m./sec. in winter is obtained from balloon measurements.

According to Johnson's measurements [5] by smoke-shells (Southern England) there is, at 30 km., a mean flow of 37 m./sec. from a westerly quadrant in winter and 12 m./sec. from an easterly quadrant in summer. An easterly wind of 30 m./sec. in summer has also been reported by Gutenberg [6] at New Mexico, U.S.A.

For the higher regions wind data are obtained from movements of noctilucent clouds [Chapter VI, Sec. 14(*b*)], from rocket data, from distortion of long-lived meteor trails and from motions of ionospheric 'clouds'.

From a 13-year series of observations on noctilucent clouds and on meteor trails carried out at Sonneberg Observatory (50°N.), Hoffmeister [7] obtained the following results. Up to the 80 km. height easterly wind prevails. Between 90 and 100 km., wind from westerly quadrant predominates, though, rather irregularly. In the region of 120 km. wind from the south-west quadrant prevails in summer, the mean speed being 50m./sec (with a maximum of 100 m./sec.) and from south-westerly or northern quadrant in winter with speed up to 270 m./sec. (mean south-westerly speed

65 m./sec. and mean northerly speed 89 m./sec.). A few observations carried out in South Africa at 23°S. lat., however, showed winds from all directions.

For the 100 km. region and also for the higher regions, up to the height of the F-layer, radio observations on 'clouds' of high ionization density indicate the presence of motions which may be regarded as winds rather than movements due to tidal pressure waves [8, 9, 9*a*] (Chapter VI, Sec. 14).

In Fig. 3, besides the wind velocities mentioned above, the meridional temperature gradients are also shown on the right.

Electric current system in the upper atmosphere.—The conducting ionized regions also participate in the tidal-cum-thermal motions described above. As the conducting regions move, cutting the magnetic lines of force of the earth, E.M.F.'s are developed which produce worldwide electric current systems (Chapter VII, Sec. 5). These current systems produce the so-called quiet day variations in the values of terrestrial magnetic elements as recorded at magnetic observatories. The current systems are, however, pronounced only in the lower ionized regions (the D-region and the E-region) where the mean free paths of the electrons are small. In the upper F-region, where the mean free paths are large, the effect of the tidal and/or thermal forces on the motions of the ions and electrons is more complicated due to the presence of the terrestrial magnetic field. They are responsible for most of the anomalous behaviours of the F-region of the ionosphere (*vide supra*).

Night air-glow.—The upper atmospheric regions are found to glow faintly at night (Chapter X, Sec. 4). Spectroscopic examination of this 'air-glow', as it is called, shows that it consists of (besides a continuous part) line radiations from O-atoms (red and green) and Na-atoms (yellow), bands due to N_2-molecule (red and blue-violet), OH-radical (red and infra-red) and O_2-molecule (blue and ultraviolet). The source of energy necessary for the excitations of the atoms and molecules is believed to be the solar radiant energy stored during daytime in the dissociated O-atoms and/or in the ionized products (electrons and ions) in the upper atmospheric regions. Many possible processes of the excitations have been suggested and examined but none of them appears to be entirely satisfactory. (For example, it has been suggested that the atomic oxygen lines and N_2-bands may be emitted in course of mutual neutralization of O^- and N_2^+ ions.) The heights of the emitting 'layers' (there may be different emitting layers for the different emissions) have been measured. The measurements are made difficult by the fact that the intensity of light is not uniform over the whole of the observed region. Some, at least, of the emitting layers appear to be identifiable with the ionospheric layers. Thus, the height of the region emitting the sodium and the oxygen lines is about 250 km., i.e. the height of the F-layer. It is also believed that part of the air-glow may be due to impact of extra-terrestrial particles with the upper atmospheric gas particles.

The fringe region of the atmosphere—the exosphere.—At the 'top' of the atmosphere the atoms and molecules have long free paths specially in the outward radial directions. The atoms and molecules in this region receiving

impacts from the air particles from the denser regions below rise and fall describing enormous orbits. (Chapter I, Sec. 5(*c*), Fig. 2). The region where the particles describe such orbits is called the *fringe region* of the atmosphere or the *exosphere.* The base of the exosphere is estimated to lie between 300 km. and 700 km. and it merges into the interplanetary space at a height of 2,000–2,500 km.

To explain the phenomena of zodiacal light it has been proposed by more than one author that the atmosphere, in its highest regions, is not symmetrically disposed round the earth. Towards the sun its extension is comparatively small; but, in the direction away from the sun it extends as a 'tail' to great lengths—to many thousands of kilometres. (See Chapter X, Sec. 2, Fig. 10.)

Atmospheric particles as are light may acquire velocity high enough to overcome the gravitational pull of the earth and escape from the atmosphere altogether. It is generally supposed that helium is thus escaping steadily from the terrestrial atmosphere. Otherwise, the helium content of the atmosphere would have been much greater due to accumulation through the geological ages [Chapter I, Sec. 5(*b*)].

Solar control.—The physical state of the upper atmospheric regions is under strong solar control.

Thus, the ionospheric conditions vary not only with the hour of the day and the season of the year, but also with the epoch of the 11-year solar cycle. (It is to be noted that the conditions of neither the troposphere nor of the ozonosphere show any association with the solar cycle variation.)

The control is exercised through the solar ultraviolet light radiation (the whole of it below 3000 A—except perhaps a short range of wavelengths at 1900 A—is entirely absorbed in the upper atmosphere), and through solar corpuscular radiation—both charged and uncharged.

The solar ultraviolet radiation consists not only of black-body radiation (as usually supposed) but also of radiation from solar flares and from the inner corona.

As a result of the ultraviolet light absorption there are allotropic modification (ozone production), ionization (formation of the ionospheric regions) and dissociation ($O_2 \rightarrow O+O$) of the upper atmospheric constituent gases.

Of the corpuscular radiation, the corpuscles which are charged concentrate round the regions of magnetic poles and are responsible for the production of magnetic disturbances and of the auroral displays. Difficulties are, however, encountered if it is assumed that particles of one sign only are travelling from the sun to the earth (Chapter IX, Sec. 1). Hence, it is believed that charged corpuscles of both signs are emitted from the sun, in the form of beams which are electrically neutral as a whole, specially during periods of high solar activity. The time taken by the particles to reach the earth from the sun is of the order of a day. Direct spectroscopic evidences have been obtained of the entry of hydrogen and calcium ions in the upper atmosphere in the polar regions [Chapter IX, Sec. 2(*d*)].

There are evidences (e.g. erratic fluctuations of the night air-glow intensity) which show that the upper atmospheric regions are also bombarded by neutral particles. These may either be of solar origin or may be existing in the interplanetary space and are swept into the terrestrial atmosphere as the earth moves in its orbit.

2. SOME UNSOLVED PROBLEMS

From the above summary it is evident that notwithstanding the many difficulties, a considerable amount of knowledge of the physical state of the upper atmosphere has been gained by the various methods of direct and indirect attack. There are still, however, large gaps in our knowledge. This is due not only to the inaccessibility of the regions concerned but also to the inadequacy of our knowledge of many fundamental laboratory data of the atmospheric constituents. But, perhaps, the greatest obstacle has been our imperfect knowledge regarding the nature of the extreme ultra-violet radiation from the sun. It is, however, hoped that this gap will soon be filled up by the rocket explorations.

In what follows we will attempt a rapid survey of the many upper atmospheric problems that are still awaiting solution.

Distributions of atomic oxygen and atomic nitrogen.—Many theoretical estimates have been made of the relative distributions of molecular and atomic oxygen in the upper atmosphere (Chapter V, Sec. 2). But, on account of the many uncertainties involved, specially temperature distribution and intensity of the dissociating radiation, the problem is not yet completely solved. (See, for example, ref. [9*b*].) Accurate knowledge of the distribution of O_2 with height in the transition region is very important, because, according to more than one author, the *E*-layer of the ionosphere is formed by ionization of O_2 in this region [Chapter VI, Sec. 10(*d*)].

Existence of nitrogen also in the atomic state in the upper atmosphere is now proved by spectroscopic evidence (Chapter V, Sec. 3). But, little is known about its relative abundance and geographic distribution. There is also difference of opinion about the mode of its production—whether it is by dissociative recombination of the N_2^+ ion or by pre-dissociation of N_2 in the range λ 1250–1150 or by direct bombardment of N_2 molecules by solar corpuscles. For the last named process the dissociative action will be confined more or less in the polar regions.

Height of diffusive equilibrium.—It is important to consider at what height above 70 km. diffusive separation becomes important, because, this will determine the relative abundance of molecular nitrogen and of atomic oxygen and atomic nitrogen in the highest regions of the atmosphere. The solution of the problem is again rendered difficult by the uncertainty regarding the temperature distribution in and above the region of dissociation of O_2. Thus, according to some estimates, the level of diffusive separation is round 300 km. (Chapter I, Sec. 4). This would mean that the highest regions of the atmosphere should consist almost entirely of atomic oxygen. But in the region of sunlit auroras, even at the height of 1,000 km., the

intensity of the first negative bands of nitrogen (N_2^+) is comparable with that of the red oxygen lines. The question, how nitrogen molecules which are nearly twice as heavy as the oxygen atoms reach such high levels against the forces of diffusive separation, is still ill-understood.

Ozonosphere.—In the ozonosphere, the modes of production and destruction of ozone and the ozone equilibrium resulting therefrom, specially, in the polar latitudes, require further clarification. Rocket experiments indicate that the intensity of the ultraviolet solar radiation in the region of ozone absorption is significantly below that assumed from 6000°K black body radiation (Chapter XIII). If this result is confirmed, the theoretical calculations on the ozone distribution that have been made (Chapter IV, Sec. 7) will have to be revised.

Interesting associations between weather conditions in the troposphere and ozone content in the stratosphere have been obtained (Chapter IV, Sec. 7). But, the ozone observations are still confined to a few stations only. A world net-work of ozone observatories, making regular records with the Dobson type of apparatus is extremely desirable. Such studies, if made in conjunction with abnormal sound propagation study and systematic meteor study, will enable one to form a world picture of ozone temperature distribution in the middle atmosphere and its diurnal and seasonal variations if any.

Temperature distribution.—The available data of the probable height distribution of temperature in the lower and the middle atmospheres and in the ionospheric regions have been collected in the preceding section. But our knowledge regarding the last two is still far from complete. It is hoped that future rocket explorations will yield more accurate data for the middle atmospheric regions, particularly for the 70–90 km. region where, due to the absence of any absorbing gas, there is a temperature drop. For the ionospheric regions a fruitful way of increasing the knowledge will be accurate measurement of the collisional frequency of electrons. Reliable knowledge of the temperature profile over stations widely distributed in latitude is very necessary to give information about the trend of diurnal variation of temperature as also latitudinal temperature field. A knowledge of the latter is essential for the study of world-wide wind circulations in the upper atmospheric regions.

Tides in the upper atmosphere.—Thanks to the works of Taylor and Pekeris (Chapter II, Sec. 4) the long-standing problem of semi-diurnal barometric oscillations, and, along with it, the difficulty of the dynamo theory of quiet day (solar) magnetic variations appear to have been solved [Chapter VII, Sec. 3(*c*)]. It is also satisfactory that mathematical treatment of solar thermal excitation of oscillations [9*c*], following the same lines as that given by Pekeris for the gravitational oscillations (for rotating earth and arbitrary temperature variation with height) has shown that thermal and gravitational contributions are of the same order as concluded earlier by Chapman. More detailed discussion of the application of this interesting work is desirable.

The lunar semi-diurnal barometric oscillations and the corresponding magnetic variations have still their puzzles. Regarding the former there is the asymmetrical distribution of the phase in time with respect to the season and also about the equator. Little is known about the origin of these asymmetries, though attempt has been made to explain them 'by the difference in resistance to tidal flow of air by surface features of the two hemispheres, the principal effective factors being the distribution of mountain ranges and the relative proportion of land and water' [10].

Ionosphere.—Exploration of the ionized regions of the upper atmosphere by the powerful 'pulse' technique has greatly advanced our knowledge of the ionosphere. It is highly satisfactory that many of the ionospheric stations established during the war are continuing their observations and that some new stations have also been established. By a synoptic study of the records of these observatories one can now form a world-picture of the distribution of ionization and its diurnal, seasonal and solar cycle variations and can even make predictions regarding its future behaviour. But still the ionosphere is full of mystery. One may, for instance, mention the comparatively frequent occurrences of localized regions of high ionization at low levels (sporadic D) in the tropical and subtropical latitudes. These ionizations are reported to have maximum intensity round Ceylon (known as *Ceylon-effect*). Very little is known about the origin of these ionizations. Again, the wavelength of the solar radiation causing radio fade-out, as also the atmospheric constituent ionized, are not known with certainty. The same remark applies to the height of the region ionized in which the enhanced S_q-type of current system flows during the fade-out (Chapter VII, Sec. 7).

Very plausible hypotheses have been put forward regarding the production of the different ionospheric regions [Chapter VI, Sec. 10(*d*)]. But, it is safe to say that the last words on the subject have not yet been spoken. For example, there are two views regarding the cause of the bifurcation of the F_1-region to F_1 and F_2 [Chapter VI, Sec. 10(*d*, ii)]. One is that it is an effect of the rapid decrease of the recombination coefficient with height and the other is that it is due to tidal effects. It is probable that both the causes are operative. But, to what degree each of these contributes to the phenomenon is still a matter of investigation.

The E-region, usually believed to be of the simple Chapman type, is now known to possess a fine structure. (Chapter VI, Sec. 4*d*. See also Ref. 10*a*.) Little is known of the origin of such structure.

The sporadic E-echoes are now believed to be caused by intrusion of localized regions or 'clouds' of high ionization density into the body of the E-layer [Chapter VI, Sec. 13(*b*)]. But, the mechanism of the reflection is not well understood. Whether it is a case of reflection, or is a scattering effect of the irregularly distributed ionospheric clouds is still debatable. The origin of these 'clouds' of ionization is also not known with certainty. Part, at least, of the E_s-ionization is to be ascribed to meteor-ionization. But, there is clear evidence that meteor-ionization does not produce all of them. 'Run-away' electrons from thunderclouds in the

subtropical regions and solar corpuscles as produce auroras and magnetic disturbances in the high latitude have also been suggested as producing extra ionization near the E-region. It is possible that all the three agencies contribute to the E_s-phenomena in different degrees in the different geographic latitudes.

In regard to the radar-echoes as obtained from meteor trails an outstanding problem is the long duration of the echoes [Chapter IIIB, Sec. 5(d)]. What is the mechanism by which sufficient electron concentration is maintained in the trail, for periods more than a minute, against the forces of diffusion? An analogous problem is the high value of night-time E-ionization. How is this high value maintained when the ionizing solar rays are not acting?

Many of the known anomalous behaviours of the F_2-region can be traced to the long free path tidal motions, of the ions and electrons under the influence of the geomagnetic field. It appears, however, that motions due to wind circulations, as developed by temperature gradients, also contribute to some, at least, of the anomalous behaviours, e.g. the so-called longitude effect [4b]. Comparative study of the two causes—winds and tides—contributing to the ionospheric anomalies is very desirable.

It is satisfactory that the intensity of the terrestrial magnetic field in the ionospheric regions as measured from the splitting of radio-echoes agrees with that calculated from the inverse cube law [Chapter VI, Sec. 4(c)]. However, unexpected seasonal variations in the intensity of the apparent magnetic field thus calculated have been reported from more than one ionospheric station. The phenomenon is certainly intriguing and demands fuller study.

Effects of tidal-cum-heating motions in the high regions of the upper atmosphere in the presence of the terrestrial magnetic field appear to explain some of the anomalous behaviours of the F_2-region. A theory—the so-called electro-dynamic theory—has been developed to explain tidal actions [Chapter VI, Sec. 14(c)]. But, the theory is still in the formative stage and requires much further investigation, both theoretical and experimental for its full development.

Some data regarding collisional frequencies in the different ionospheric regions have already been obtained. But they are still very meagre and not very accurate. More careful determination in the different latitudes is needed to give reliable information about temperature distributions in the upper atmospheric regions.

The high value—several orders higher than the theoretically computed value—of the recombination coefficient of ions and electrons in the ionospheric regions is another problem which requires further investigation. The hypothesis of effective-recombination coefficient appeared to have offered a satisfactory explanation of such high value [Chapter VI, Sec. 12(c)]. Further, the identification of the nocturnal F_2-region with the night airglow layer emitting the N_2-bands and O-lines helped to unify the effective recombination hypothesis with the emission processes of these lines and

bands. However, contemporary work questions the soundness of the effective-recombination hypothesis. In the theory of tides in the F_2-region referred to above, if account is taken of the electrodynamic forces developed as a result of the tidal motions, then a term appears in the expression of the recombination equation which becomes important in this region. The value of the recombination coefficient, as computed taking into account the contribution by this term, agrees with the observed value. It is desirable that a closer comparative study of the effective recombination hypothesis and the electrodynamic hypothesis be made to estimate the relative importance of the two in bringing agreement between observed and theoretically computed values.

Associations between ionospheric conditions, on the one hand, and barometric conditions at the ground, on the other, have been reported from time to time (Chapter VI, Sec. 16). If these associations be real, then they would be valuable guides to weather prediction. More observations on a world-wide scale on the reality of such associations are needed.

The complicated latitude and longitude variations of F_2-region ionization involving both geographical and geomagnetic equators raises the important question of defining an 'ionospheric equator'. Only preliminary works in this line have been done.

Our progress of knowledge on the ionosphere depends to a great extent on the availability of world-wide ionospheric data. Unfortunately, the ionospheric stations are not distributed uniformly over the two hemispheres; there are more stations in the northern than in the southern hemisphere. Consequent lack of data is a great handicap to the solving of many ionospheric problems. More stations in the southern latitudes are very necessary. It is possible that a sea expedition might secure useful ionospheric data from places which are scientifically important, but for which data are lacking.

Terrestrial magnetic variations.—An intriguing problem in connection with the S_q and the L current systems which produce the quiet day solar and lunar magnetic variations is the relative changes in their intensities with the season and with the solar cycle (Chapter VII, Sec. 4).

Both S_q and L increase greatly in intensity in June in the northern hemisphere. But, the increase in the L-current system is larger than that for the S_q-system.

From the above one would expect that the solar cycle variation of L would be larger than that of S_q. But, the reverse is the case. While the variation of S_q is about 50% that of L seems only 20%. The reason for these curious anomalies is not known and requires investigation. It is probable that the S_q- and the L-current systems flow at different levels. One should also remember that the modes of solar and lunar tidal oscillations (which are the ultimate cause of S_q and L) might be different (Chapter II, Sec. 4).

Amongst other anomalies in the S_q- and L-variations which are still largely unexplained, one may mention the quite large day to day changes

in the corresponding current systems (the positions of the current foci altering by as much as 15°) and the considerable variations in the value of H—nearly twice as large as at other places—observed at Huancayo situated on the magnetic equator [Chapter VII, Sec. 4(*a*)]. S_q augmentation at this station is also abnormally large during solar flares. It is, however, remarkable that while S_q and L variations in H are abnormally large, the geomagnetic disturbance effects (i.e. S_D, D_{st} and D_m) are quite normal. To explain the large range of H variation it has been postulated that there is a concentrated upper atmospheric eastward electric current system—*electrojet*—over the magnetic equator (the intensity of which rises and falls) superposed on the normal S_q current system [11]. (Investigations have shown that the electrojet exists over Africa and India.) Attempt has been made to explain the *electrojet* as due to existence of a narrow region of very high conductivity, along the line of zero dip value, near the E-region as a result of polarization by Hall current [11*a*]. But, clearly, the subject requires further study.

Regarding the current systems producing the disturbance type of variations, the origin of neither the Schmidt's wandering vortices, nor of the clockwise and anti-clockwise rotation of the total force vector during magnetic storms (as discovered by Sangster) is well understood [Chapter VII, Sec. 6(*c*)]. These phenomena may probably be connected with non-uniform conductivity of the current carrying regions, but they need further investigations.

Night air-glow.—A large amount of work has been done on the night air-glow (Chapter X, Sec. 4). But our knowledge about the phenomenon is still far from complete. A broad question has been raised: Whether all the energy of excitation of the glow is derived from the stored solar energy in the upper atmosphere (stored during day-time in course of ionizing and dissociating the upper atmospheric gases) or, at least, a part of it is due to impact of extra-terrestrial particles.

A simple process of excitation of the O-lines and the N_2-bands has been proposed by more than one author. But the efficiency of the process has been questioned. Further, some workers have begun to cast doubt on the identification of the Vegard-Kaplan bands and also of the first positive bands of N_2—the latter on account of the presence of OH-bands in the vicinity.

It is satisfactory that the red radiation at 6580 A and the strong infra-red band at 10,400 A have been identified as OH-bands. But, the mode of excitation of the bands is not known with certainty. One of the suggestions made is that they are emitted during, rather than after, the production of OH.

The continuous spectrum in the background of the lines and bands of the night air-glow appears to have received inadequate attention. More investigation about its intensity distribution and the origin is necessary.

A problem of great interest and importance is the measurement of the heights of the emitting layers (Chapter X, Sec. 6). The measurement is

rendered difficult by the non-uniform intensity distribution of the emitting layers. However, it is satisfactory that fast-reading photo-electric photometers have been devised by which it is possible to scan the whole sky in half of an hour's time. Though it appears that the different layers have different heights, there is a main layer near the F_2-region. It is highly desirable that the height measurements be carried out systematically in different parts of the world and relations, if any, of the luminescent layer with the F_2-layer be studied.

Of the other lights of the night sky the mystery of the zodiacal light is still not completely solved [Chapter X, Sec. 2(*c*)]. It appears, however, that there has been some rapprochement between the two rival theories—the atmospheric theory and the planetary theory. While the main phenomena of zodiacal light are attributed to scattering of sunlight by interplanetary dust particles, a substantial portion of the light is due to fluorescent scattering of the terrestrial atmosphere which is assumed to be greatly extended in the direction opposite to the sun. However, these hypotheses are still in the formative stage; there are many gaps and unexplained details. Further observations, specially photometric measurements from different parts of the world all the year round, are necessary to help in building up a complete theory.

Auroras and magnetic storms.—The problems of aurora and of magnetic storms are closely linked up. The view long held that the primary cause of these phenomena is incidence of fast charged particles emanated from the sun has now been confirmed by spectroscopic evidences [Chapter IX, Sec. 2(*d*)]. It is also satisfactory that the neutral solar corpuscular beam theory (proposed to meet the fundamental Schusterian difficulty) is making headway. The motions of the positive and negative particles become differentiated by the terrestrial magnetic field and, with plausible assumptions, their incidence on the upper atmosphere may explain the auroral phenomena. However, much work, both theoretical and observational, remains to be done to give complete shape to the theory.

Of the auroras we do not yet know how they assume the various forms—long streamers, arcs, draperies, etc. Nor is it known how auroral form of one type changes over to another type, e.g., a homogeneous arc gradually changing over to a drapery-shaped arc and then transforming into draperies and ray bundles. The distribution of intensity with height is also still imperfectly understood. Of the excitation of the auroral spectra the only process regarding which there is general agreement is that the first negative bands of N_2^+ are excited by direct bombardment of the N_2 molecules by the impinging corpuscles. Regarding the modes of excitation of the forbidden lines of O, of N and weak forbidden bands of N_2 there is diversity of opinion. The identification of strong permitted lines of O and N has added to the list of difficulties.

Lack of laboratory data.—As regards laboratory data, the lack of which is still standing in the way of the advancement of our knowledge of the upper atmosphere, one may mention the following: More complete data regarding

the electronic spectra and the absorption coefficients of N_2, O_2 and O in the extreme ultraviolet region are very necessary, because, the ionizations in the upper atmosphere are controlled by the absorptions by these gases in the extreme ultraviolet.

The value of electron-ion recombination coefficient and its variation with temperature are parameters of immense importance in ionospheric studies. It is satisfactory that laboratory measurements of the recombination coefficients have already been undertaken and specially with the aid of the powerful microwave technique [Chapter VI, Sec. 12(*d*)]. It is hoped that the investigations will help to solve many of the puzzles of recombination phenomena in the ionospheric regions.

Closer laboratory investigation of the 'rotational' temperature is very necessary to solve the puzzle of the measured 'rotational' temperature in the auroras (deduced from band spectra) being consistently lower than the gas temperature deduced from other sources. There are many instances of laboratory experiments in which the rotational temperature is *higher* than the gas temperature, but, instances of the 'rotational' temperature being lower are rather rare [12, 13]. The laboratory studies would provide valuable information regarding conditions of excitation in the upper atmospheric regions.

Theoretical study of the many quantal processes that occur in the upper atmospheric regions, e.g., collision, absorption, the many dissociative processes—photo-dissociation, dissociative recombination and others is extremely important. As an example of the importance of such study one may mention about the values of the collision frequencies of electrons with atmospheric particles as deduced from ionospheric observations [Chapter VI, Sec. 7(*b*)]. One is tempted to apply gas kinetic relations and deduce the molecular densities in the ionosphere regions from these values. This, however, is hardly justifiable, because, the cross-sections of atoms and molecule for low velocity electrons (a few tenths of electron volt corresponding to some 1000°K) may be widely different from the gas kinetic cross-section. Attention may specially be drawn, in this connection, to the case of electron-atomic oxygen collisions, which are of great importance in the ionospheric regions. A number of theoretical studies have been made [14, 15], but the values of the cross-sections obtained are far from concordant.

More theoretical and experimental data on the collision of high velocity neutral particles with solid surfaces will be very helpful in understanding the phenomenon of evaporation of meteors rushing through the atmosphere [Chapter III B, Sec. 4(*c*)].

APPENDIX

1. ILLUMINATION OF THE UPPER ATMOSPHERE BY SOLAR RAYS

Upper atmospheric phenomena such as ionization and dissociation of the constituent gases are controlled by the incidence of solar radiation. As such, it becomes necessary to know the hours at which the sun's rays strike or disappear from different levels of the atmosphere during the pre- or post-sunrise periods at the surface of the earth. The methods of calculating these hours, as also the results obtained for a few typical cases, are discussed in this section.

Hours of sunrise and sunset at different atmospheric levels.—The shadow of the earth in space is of cylindrical shape. The height H at which the cylinder cuts the zenith is given by

$$H = a\left(\frac{1}{\cos\theta} - 1\right) \quad .. \quad .. \quad .. \quad (1)$$

where a is the earth radius and θ is the angle of depression of the sun below the horizon. This is evident from Fig. 1.

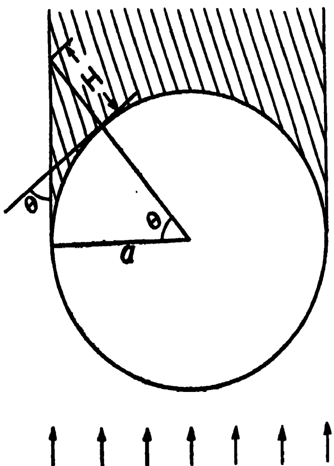


Fig. 1. The shadow of the earth in space is cylindrical. If θ is the angle of depression of the sun below the horizon, the height H at which the cylinder cuts the zenith is given by $a\,[(1/\cos\theta)-1]$.

The hour angle h of the sun at its rising at height H may be obtained as follows. We have,

$$\cos Z = \sin\phi \sin\delta + \cos\phi \cos\delta \cos h \quad (2)$$

where

Z = zenith distance of the sun,

δ = its declination,

ϕ = latitude of the place of observation.

From Fig. 1, $Z = 90° + \theta$.

In the above the effect of refraction which accelerates the time of rising has not been taken into account. Thus, allowing 34′ for horizontal refraction and 16′ for semi-diameter, the true zenith distance of the sun's centre, when it rises at a height H above the ground, is given by

$$Z = 90° + \theta + 50' \quad .. \quad .. \quad (3)$$

Moreover, at actual noon the apparent time is 12^h and the mean time is $12^h + \epsilon$, where ϵ is the equation of time. Hence the mean time t of sunrise at a height H is given by,

$$t = 12^h + \epsilon - h.$$

Similarly, the hour of sunset at H is given by

$$t = 12^h + \epsilon + h.$$

Knowing the values of δ and ϵ from Nautical Almanac the hours of sunrise and sunset at different atmospheric levels may be calculated with the help of Eqs. (1) and (2). In Figs. 2 (a) and (b) curves delineating the variation of the hour of sunrise with height for the whole year, at intervals of about a fortnight, are given, after Ghosh [1] for the latitude (22° 32′ 48″ N.) of Calcutta. In Figs. 3 and 4 the curves give the hours of sunrise at different latitudes on the surface of the earth and also at heights 50, 100, 250, 500 and 1,000 km. above the surface of the earth on June 21 (summer solstice) and on March 21 (vernal equinox) [2]. Fig. 5 gives the heights above which, in the latitudes 40° to 90°, the atmosphere is illuminated by solar rays at midnight in course of the year [2]. It will be noticed that even at the middle latitude of 55° the atmosphere above 100 km. is illuminated at midnight during high summer. It is also noteworthy how the height at which illumination commences, varies rapidly with the season of the year particularly in high latitudes.

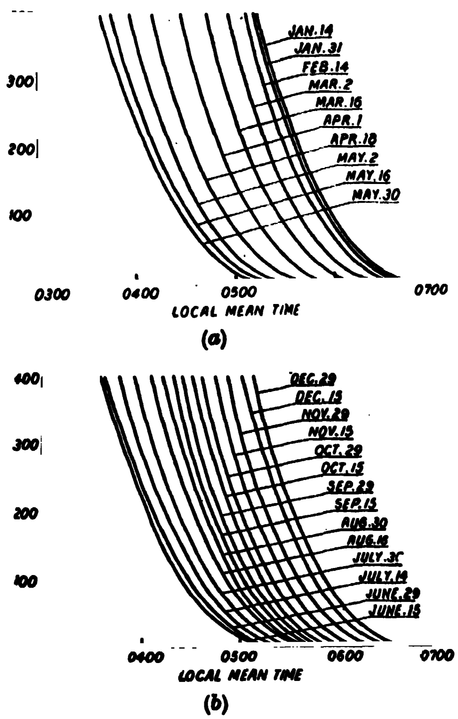


FIG. 2. Curves delineating the variation of the hour of sunrise with height at intervals of about a fortnight for the latitude of Calcutta.

Hour at which solar rays strike a particular atmospheric level by grazing a surface concentric with the earth.—For certain purposes it is necessary to know the hour at which the rays of the sun strike a particular region of the atmosphere at a height H by grazing a concentric surface d km. above that of the earth [Chapter VI, Sec. 12(e)]. It will be seen from Fig. 6 that when the rays of the sun strike at B after grazing a concentric surface d km. above that of the earth, it is sunrise at A. From similar triangles BNO and AMO

$$\frac{OA}{OM} = \frac{OB}{ON}$$

or

$$x = a\frac{H-d}{a+d},$$

where $BL = H$, $AL = x$, $MN = d$, $OL = a =$ radius of the earth. Thus from a knowledge of H and d, x can be calculated. It will be seen that for a given value of d, the relation between x and H is linear. The hour of sunrise at A and hence the hour when the rays of the sun strike at B after grazing the surface under consideration, can be easily calculated. One point should be mentioned in this connection. When the value of d is large so that the rays of the sun in order to strike at B pass through a portion of the atmosphere of low density, the effect of horizontal refraction is negligible. In that case, the zenith distance of the sun's centre is given by

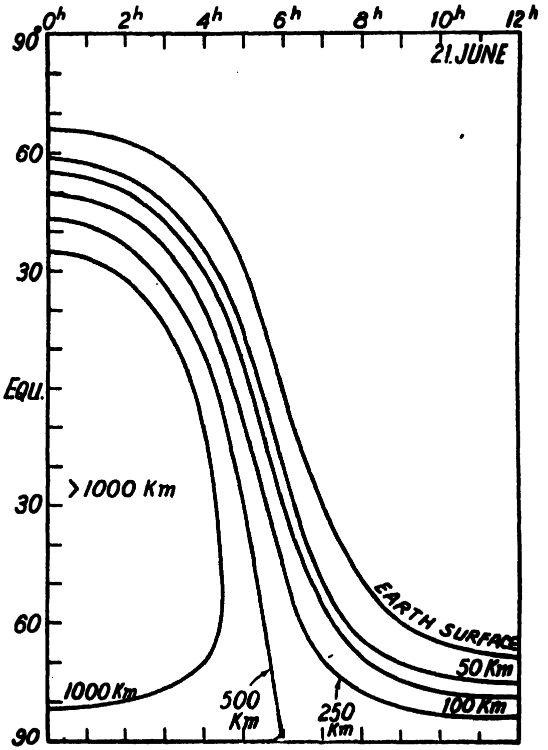


FIG. 3. The curves give the hours of sunrise at different latitudes on the surface of the earth and also at heights 50, 100, 250, 500 and 1,000 km. above the surface on June 21. Ordinate—latitude in degrees. Abscissa—meridians indicated by local time. (After Bartels.)

$$Z = 90° + \theta + 16'.$$

Amounts of solar energy received per day at different latitudes in various seasons.—It is interesting to consider the amount of solar energy received in course of a day at different

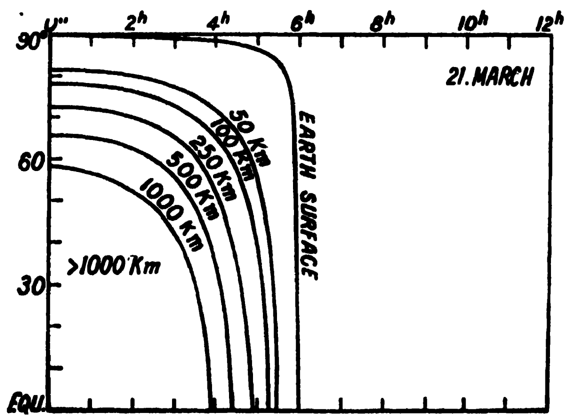


FIG. 4. Same as Fig. 3 for March 21. (After Bartels.)

latitudes in the various seasons of the year. The solar rays are strongest in the equatorial region on account of the fact that they are incident here

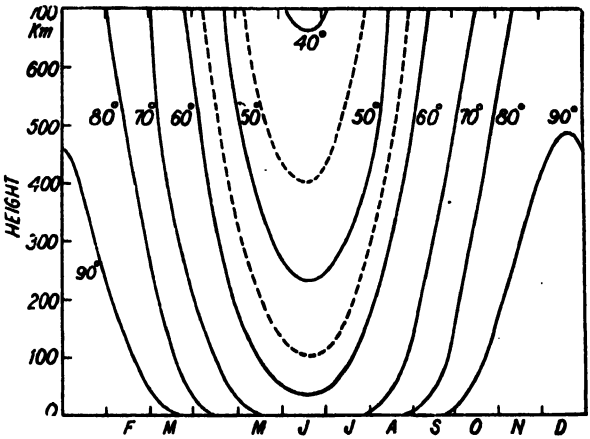


FIG. 5. Heights above which in the latitudes 40° to 90° the atmosphere is illuminated by solar rays at midnight in course of the year. It is to be noticed that the height in question varies rapidly with season particularly in high latitudes. (After Bartels.)

nearly normally. If, however, the total amount of solar energy received in course of a day is calculated for different latitudes it is found that the amount may be greater in high than in low latitudes in particular seasons of the year. This is because the decrease in intensity due to the obliquity of the rays in the high latitude is more than compensated by the longer duration of the daylight hours. For instance, it may be shown that on a summer day the polar region receives more than 30% of the solar rays in an equatorial region. Fig. 7, after Gessler [3], shows how the total solar energy received by a horizontal sq. cm. surface in course of a day, varies with the latitude for the dates —February 4, March 21, May 6, June 21, September 23 and December 22.

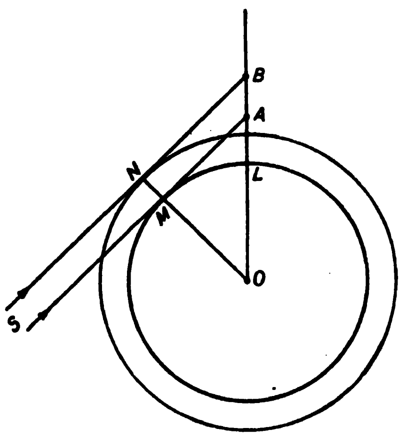


FIG. 6. The sun's rays strike *B* by grazing a concentric surface *d* km. (= *NM*) above the surface of the earth.

The scale of the ordinate on the left is given in the so-called *equator-hour* which is the amount of energy received by a horizontal surface of one sq. cm. at the equator from normally incident solar rays in one hour. The scale of the ordinate on the right is obtained by multiplying the equator-hour by the solar constant (1·93 cal./cm.^{2} min. = 1·35 $\times$ 10^6 erg./cm.^{2} sec.). It is to be noted that the amounts of energy shown are those received

outside the atmosphere, i.e., the absorption by the atmosphere is not taken into account.

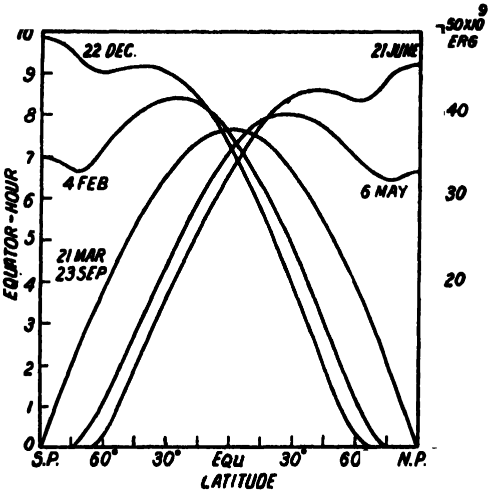


FIG. 7. Illustrating how the total solar energy received in course of a day per sq. cm. surface, horizontally held, varies with latitude. For explanations of the ordinate scales see text. (After Gessler.)

2. SPECTROSCOPIC NOTES

In course of our discussions on the action of solar ultraviolet rays on the upper atmosphere we had had many occasions to refer to the spectroscopic states of the constituent gases. For convenience the relevant data of nitrogen and oxygen—the commonest constituents—will be summarized and classified in this section. Attempt will first be made, however, to explain in a general way the spectroscopic terminology. This will serve as ready reference to those not versed in spectroscopy. For fuller details the reader must consult specialized treatises, e.g. *Atomic Spectra and Atomic Structure* (Prentice-Hall, 1937) and *Molecular Spectra and Molecular Structure* (Prentice-Hall, 1939), by G. Herzberg.

(a) Atomic Spectra

The fundamental discovery that an atom (or a molecule) cannot exist in state of any energy, but only in one of a set of discrete energy states was the starting point of the quantum theory.

In many simple cases these energies (their relative values omitting an arbitrary constant) can be determined directly by electron bombardment experiment (e.g. Franck and Hertz's experiments). More generally, they are deduced from spectroscopic data—the frequencies of the lines and bands emitted. Because, according to Bohr's frequency condition when the atom executes a quantum jump from a state of higher energy E_n to one of

lower energy E_m it emits radiation of frequency ν_{nm} given by the relation $E_n - E_m = h\nu_{nm}$, where h is the Planck constant.

Besides the discrete energy states the atom has other sets of discrete states defined by the values of the angular momentum, the direction of the angular momentum vector in space and the spin axis of the electron. All these different states contribute to the total energy of the atom. In what follows we will only formulate and explain the so-called quantum conditions by which these discrete sets of the states are selected and leave their derivations to specialized treatises.

To understand the quantum conditions it is best to consider the simplest atomic structure, namely an electron moving in the field of a heavy, central positive point charge. The hydrogen atom conforms to this ideally. For other cases, when a number of electrons are present, one can select for consideration one of the outer electrons which takes part in the spectral processes and regard the action of the remaining electrons, together with the nuclear positive charge as that due to a positive 'core' giving rise to a centrally symmetrical field. The representation is, of course, approximate but simulates the ideal case the further the electron under consideration is removed from the 'core'.

(i) *Energy* (*total quantum number*).—The discrete set of energy states is given by $E_n = f(n)$, where $n = 1, 2, 3, \ldots$. For the case of hydrogen-like atoms it is given by $E_n = -RZ^2/n^2$ where R/hc is the *Rydberg constant* $= 2\pi^2\mu e^4/ch^3$; Z—atomic number; $\mu = mM/(m+M)$; m—electronic mass; M—nuclear mass. n is called the *principal quantum number*.

For the more general case for which the 'core' is not a point charge, the energy is given approximately by $E_{n,l} = -RZ^2/(n+a)^2$ where a is the *Rydberg correction* depending on l (see *infra*). Z here is charge on the 'core' of the atom. Note that the expression for the energy is negative. This is because it has an additive constant which has been put equal to zero. The spectroscopic *term* is proportional to E_n (term $= -E/hc$). Often, the words, *term*, *energy state*, *quantum state* are used synonymously. And, *term value* signifies *energy value*.

A series of terms of the form $RZ^2/(n+a)^2$ is called a *Rydberg series* of terms.

(ii) *Angular momentum or moment of momentum* (*azimuthal quantum number*).—For classical motion in a central field an important characteristic is that the angular momentum (the vector product of the orbital momentum vector and the radius vector) is a constant of motion. In quantum mechanics the angular momentum $\boldsymbol{l}$ can take up the set of values given by $\sqrt{l(l+1)}h/2\pi$, where $l = 0, 1, 2, \ldots$. For large values of l, the magnitude of $\boldsymbol{l}$ is approximately equal to $lh/2\pi$. For $l = 1$, it is equal to $\sqrt{2}h/2\pi$.

(iii) *Space quantization* (*magnetic quantum number*).—If a direction in space be specified by a field $\boldsymbol{F}$ (magnetic or electric field however small—in the limiting case zero field) then the angular momentum vector precesses round it (Fig. 1). According to quantum mechanics the precession takes place at angles with the field direction such that the component of the

angular momentum in this direction has one of the set of discrete values given by $m_l h/2\pi$ where $m_l = 0, \pm 1, \pm 2, \ldots \pm l$. In other words, if θ be the angle between l and the specified direction, then $\cos\theta = m_l/\sqrt{l(l+1)}$ (Fig. 2). Note that the minimum value of θ is not zero, though it approaches the same for large values of l; the minimum value of θ for $l = 1$ is $\cos^{-1} 1/\sqrt{2}$. For $l = 0$, $\cos\theta$ becomes indeterminate, i.e., the distribution-in-θ of the angular momentum vector is symmetrical. In Fig. 2 space quantization for the case $l = 3$ is illustrated.

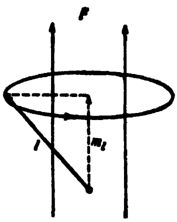


FIG. 1. The angular momentum vector l of an electron precesses round a field F, magnetic or electric.

(iv) *Spin vector* (*spin quantum number*).—Investigations on the multiplicity of spectral lines, as also on the anomalous Zeeman Effect have led to the conclusion that the electron behaves as if it possesses an angular momentum of its own, i.e., as if it spins on its own axis. Dirac's theory of electron based on quantum mechanics also leads to the same conclusion. The magnitude of the spin momentum s is given by $\sqrt{s(s+1)}h/2\pi$ where s is $\frac{1}{2}$. In a magnetic field the spin vector can only so orient itself that its component along the field has the value $m_s h/2\pi$ where m_s can only be $+\frac{1}{2}$ or $-\frac{1}{2}$. Approximately, it may be said, that the spin vector sets itself parallel or anti-parallel to the field.

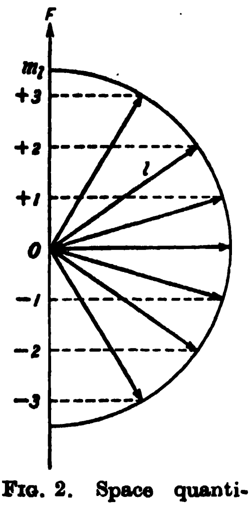


FIG. 2. Space quantization of angular momentum vector ($l = 3$) in a field F. Note that the lengths of the vectors are proportional to $\sqrt{l(l+1)}$. As such l cannot coincide with F, though, for large values of l it approaches the same.

From the above it is seen that an electron in an atom is characterized by four quantum numbers:—

(1) n—Principal quantum number; this determines the energy of the electron; $n = 1, 2, 3, \ldots$

(2) l—Azimuthal quantum number; this gives the orbital angular momentum; if the principal quantum number is equal to n, then $l = 0, 1, 2, \ldots, n-1$.

(3) m_l—Magnetic or latitudinal quantum number; this gives the orientation of l with respect to a specified direction in space; $m_l = 0, \pm 1, \pm 2 \ldots \pm l$.

(4) m_s—Spin quantum number giving the spin vector of the electron; $m_s = \pm\frac{1}{2}$.

Electrons with $l = 0, 1, 2, 3, \ldots$ are designated s, p, d, f electrons respectively. (This s is not to be confused with the spin quantum number s.)

Pauli's Principle of Exclusion—Atom building.—According to this principle *no two electrons in an atom can have the same set of the four quantum numbers* n, l, m_l, m_s. The principle gives a clear picture of the atom building process starting with the simplest one-electron atom, hydrogen, and, at the same time, explains the periodic classification of the elements. The latter is depicted in Fig. 3 after Mitra [1]. The lines connecting

the elements indicate the sequence of the building up process. It is to be noted that this follows the sequence of the increasing atomic numbers and

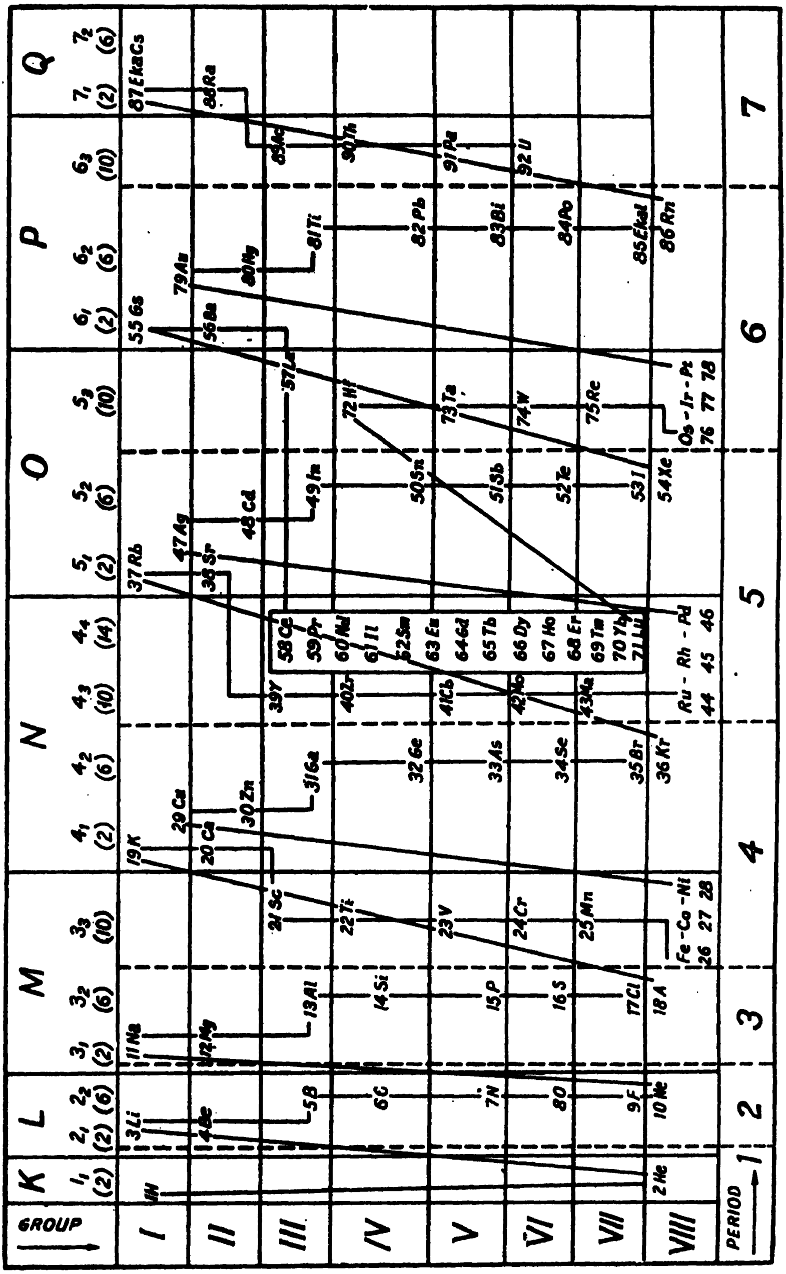


FIG. 3. Periodic chart of the elements. [After Mitra. For details, see *Phil. Mag.*, 11, 1201 (1931).]

not necessarily that of the increasing atomic weights. The figure together with Table I below showing the building up process of the first few elements, will help to make the principle clear.

It is evident that the number of electrons in the group characterized by $n = 1$ or the K-shell as it is called, can be only 2. Similarly, the number in the L-shell ($n = 2$) can be only 8. Also, the number in the ' cell' characterized by $n = 2$ and $l = 1$ can only be 6. It can easily be verified that for a shell with principal quantum number n, the number of electrons is $2n^2$ and for a 'cell' with azimuthal quantum number l, the number of electrons is $4l+2$.

TABLE I

Electron configuration of an atom up to $n = 3$ after Pauli's principle of exclusion.

	K	L				M								
n	1	2				3								
l	0	0	1			0	1			2				
	s	s	p			s	p			d				
m_l	0	0	-1	0	$+1$	0	-1	0	$+1$	-2	-1	0	$+1$	$+2$
m_s	$+\ -$	$+\ -$	$+\ -$	$+\ -$	$+\ -$	$+\ -$	$+\ -$	$+\ -$	$+\ -$	$+\ -$	$+\ -$	$+\ -$	$+\ -$	$+\ -$

As an example we give the electronic structure of the nitrogen atom. The position of nitrogen in the periodic table is the seventh (Fig. 3). Its electronic structure can therefore be represented as—$(1s)^2(2s)^2(2p)^3$. This means that in the cell characterized by $n = 1$ and $l = 0$, there are two electrons; in the cell $n = 2$ and $l = 0$ there are two electrons; in the cell $n = 2$ and $l = 1$, there are only three electrons, though this last one can hold a maximum of six *equivalent electrons* (having the same n and the same l). This cell is therefore incompletely filled in the nitrogen atom. Similarly, the oxygen atom whose place in the periodic table is the eighth has the electronic structure—$(1s)^2(2s)^2(2p)^4$.

Addition of angular momentum vectors.—When more than one electron in an atom take part in the light emission process, then the angular momenta of all of them have to be taken into account. All the electrons in the outermost cell—complete or incomplete—take part in the excitation and emission processes from an atom. Due partly to electric repulsion and partly to the magnetic moments resulting from their angular momenta there is strong interaction between the individual electrons. In such cases the individual angular momenta are added up according to quantum addition rules and the resultant vector takes up quantized positions with respect to a specified direction in space. The rule can best be illustrated

by an example. Take two angular momenta vectors with l values 1 and 2 i.e., the magnitudes of $\boldsymbol{l}$ are $\sqrt{2}h/2\pi$ and $\sqrt{6}h/2\pi$ respectively. Then the two $\boldsymbol{l}$'s, as shown in Fig. 4, are so compounded that the values of the resultant vectors ($\boldsymbol{L}$) are given by $\sqrt{3(3+1)}h/2\pi$, $\sqrt{2(2+1)}h/2\pi$ and $\sqrt{1(1+1)}h/2\pi$. (Note that the lengths of the vectors, both originals and the resultants are of the form $\sqrt{l(l+1)}h/2\pi$ but for simplicity they have been drawn in Fig. 4 proportional to l or L.) Thus the resultant $\boldsymbol{L}$ of angular momentum vectors $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ when compounded can have the following values

$$L = (l_1+l_2), (l_1+l_2-1), (l_1+l_2-2) \ldots . |l_1-l_2|.$$

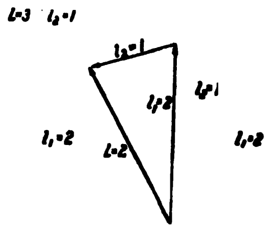


FIG. 4. Illustrating the rule of addition of $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ to give resultant orbital angular momentum $\boldsymbol{L}$. $l_1 = 2$, $l_2 = 1$. $L = 1, 2, 3$.

The individual angular momenta ($\boldsymbol{l}_i$) which are compounded to give an $\boldsymbol{L}$, precess round the latter as illustrated in Fig. 5.

Terms arising out of states for which the L values are 0, 1, 2, 3, ... are called S, P, D, F-terms respectively.

Total angular momentum J: Multiplicity of terms.—Just as the angular momenta ($\boldsymbol{l}$'s) of the individual electrons combine to give a resultant $\boldsymbol{L}$, so the spin momenta of the individual electrons combine to give a resultant spin $\boldsymbol{S}$. (This $\boldsymbol{S}$ is not to be confused with the term symbol S corresponding to $L = 0$.) Now since the spin quantum number s can only have values $\pm\frac{1}{2}$ it follows that the resultant S is zero, integral or half-integral. According to quantum mechanics the resultant $\boldsymbol{S}$ and the resultant $\boldsymbol{L}$ cannot be oriented with respect to each other in any manner but can take up only a set of discrete positions. The resultant vector $\boldsymbol{J}$, total angular momentum, has only the values $\sqrt{J(J+1)}h/2\pi$, where J is given by

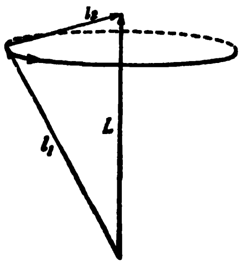


FIG. 5. The two angular momenta $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ precess about the resultant $\boldsymbol{L}$.

$$J = (L+S), (L+S-1), \ldots . . |L-S|.$$

This is illustrated in Fig. 6 for $L = 1$ and $S = 1$.

If there is an external magnetic field not too strong then each $\boldsymbol{J}$ can take up quantized positions inclined to the direction of the field and precesses round it—the so-called *Larmor precession*.

Now, the energy due to the angular momentum is much larger than that due to the spin. The different energy values due to $\boldsymbol{J}$, arising out of an $\boldsymbol{S}$ combining in different ways with the same $\boldsymbol{L}$, thus lie close to each other. This gives rise to what is called *multiplet terms*. From the rule for forming the resultant $\boldsymbol{J}$'s, it easily follows that the number of $\boldsymbol{J}$ values corresponding to a given L is equal to $(2S+1)$ when L is greater than (or equal to) S. $(2S+1)$ is called the *multiplicity* of the term. For the

case in Fig. 6 ($L = 1$, $S = 1$) the P term is a triplet and its symbol is $^3P_{2,1,0}$. The multiplicity ($2S+1 = 3$) is indicated by the upper left index. The possible J values are indicated by the lower right indices. It is to be noted that if $S > L$, the number of the possible J values, that is, the number of the split components is given by $2L+1$. However, the *multiplicity* is still indicated by $(2S+1)$. This is because the behaviour of a term depends rather on the magnitude of its additional momentum due to the spin vector S, than on the number of components into which it may be split. The S term ($L = 0$) is always single but it may have a higher multiplicity.

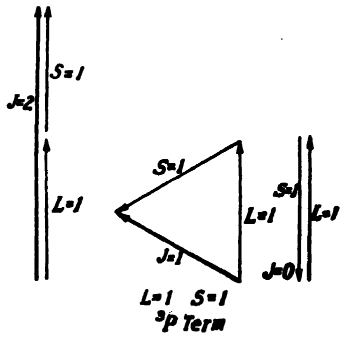


FIG. 6. Vector addition of L and S for the case $L = 1$, $S = 1$. The resultant $J = 2$, 1 and 0. L and S together precess about the resultant J (compare Fig. 5). The original L, by different combinations with S, gives rise to the triplet P term.

Selection rules: Forbidden transitions. Transitions, spontaneous or by absorption of radiation, do not occur between any two energy states of the atom. They are governed by the so-called *selection rules*. These are:—

(1) S (spin) can change by 0 or ± 1. The former holds good for a special type of coupling between the angular momentum vectors and the spin momentum vectors known as Russell-Saunders coupling.

(2) L can change by 0 or ± 1.

(3) J changes by 0 or ± 1; a change of 0 to 0 is excluded.

(4) If the terms are classified as odd or even according as the arithmetical sum of l-values of all the electrons is odd or even, then, transition may occur only between odd and even terms, not between two odd or two even terms.

Notwithstanding the above restrictions transitions do occur, under special circumstances, which are forbidden by the selection rules. Since the lines due to permitted transitions are not all of equal intensity, it is evident that even the permitted transitions have not all the same probability of occurring. The forbidden lines may thus be regarded as due to transitions, the probabilities of which are not zero but are very small compared to those of permitted transitions. Forbidden transitions are closely related to the occurrence of the metastable states (vide *infra*) and forbidden lines are observable only under special conditions [Chapter X, Sec. 4].

(*b*) Molecular spectra

Energy states—Band systems.—The different energy states of an atom arise out of its electronic states, e.g., the distances of the electrons from the core, their angular momenta and spin. In a molecule, besides the electronic states, the vibrations of the nuclei about their equilibrium position and

their rotations also contribute to the energy of the state. The total energy of the molecule may therefore be written as

$$E = E_e + E_v + E_r$$

the symbols being self-explanatory.

Since the number of energy states is thus greatly increased, the numbers of possible transitions between the different states are also enormously increased. Molecular spectra are thus more complicated than

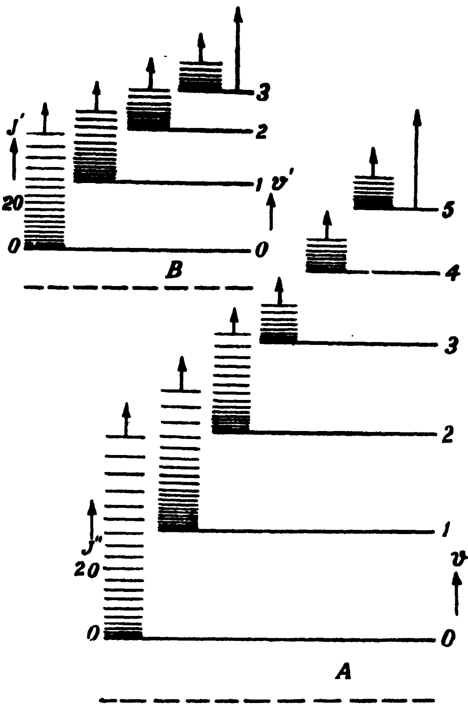


FIG. 7. Illustrating the electronic, vibrational and rotational levels of a diatomic molecule. A and B (broken lines) are the two undisturbed electronic levels. When the nuclei vibrate they give rise to vibrational levels (v', v'' = 0, 1, 2, 3, etc.). The two nuclei may also rotate round an axis perpendicular to the line joining the two. Each electronic vibrational energy level is thus accompanied by rotational energy levels J', J'' = 0, 1, 2, 3, etc.). In the diagram alternate rotational levels are omitted. An electronic transition may thus be accompanied by vibrational and rotational transitions.

the atomic spectra. In what follows we shall only consider the simplest type of molecular spectra—the so-called band-spectra of diatomic molecules. In Fig. 7 attempt is made to illustrate the vibrational and the rotational energy levels associated with two electronic states A and B (broken lines). The vibrational levels (long full lines) are designated v' and v'' and marked 0, 1, 2, etc. The rotational states (short thin lines) are designated J', J'', the alternate lines being omitted.

To understand the nature of the transitions between the various states we first recall that the electronic energy of a given electronic state depends on the internuclear distance. But, as the electrons, on account of their smaller mass, move much more rapidly than the nuclei, they take up, at any instant, during the nuclear vibrations, positions appropriate to the internuclear distance.

Let us now consider transitions between two specified vibrational levels, for example, $v' = 2$ and $v'' = 1$, belonging respectively to two electronic

states B and A as in Fig. 7. Each vibrational level is associated with rotational levels. Hence, there will be rotation-vibration transitions and the $v' \rightarrow v''$ transitions will consist of many lines, closely packed, on account of the closeness of the rotational levels. In spectroscopes of low or medium dispersive power these lines, being unresolved, merge into one another and appear as a band of continuous spectrum. (In the case under illustration this will be called a 1-2 band.) The rotation-vibration transitions from electronic state B to electronic state A will thus consist of a *system of bands*. The series of bands due to transitions between a given upper vibrational level (v') to lower levels $v'' = 0, 1, 2$, etc. is called *v''-progression*. Similarly, the series of bands due to transitions between a given lower vibrational level (v'') and the upper levels $v' = 0, 1, 2$, etc. is called *v'-progression*. It will be seen from the diagram that the differences between the pairs of levels for which Δv ($=v'-v''$) is constant, are nearly the same. Hence, the bands due to such transitions will lie relatively close to each other. Such bands are called a *sequence of bands*. Thus the bands 0–2, 1–3, 2–4, etc. form a sequence for which $\Delta v = -2$.

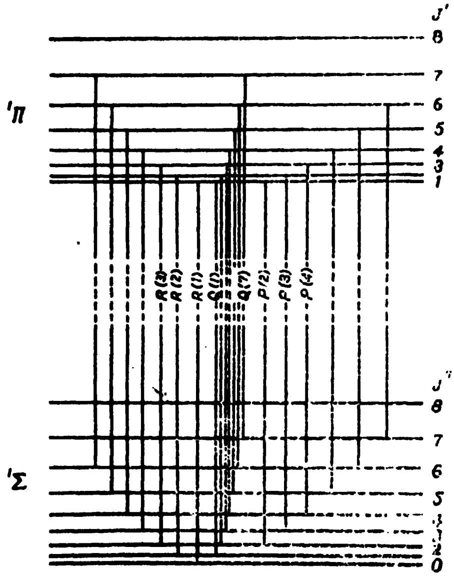


FIG. 8. Rotational energy levels of two electronic states illustrating transitions corresponding to the P, R and Q branches of a band. Note that all the three branches, P, R and Q of a band are developed only if for, at least, one of the states Λ is not zero. (If for both the electronic levels $\Lambda = 0$, then the transitions $\Delta J = 0$ are excluded, i.e., there is no Q branch.)

With higher resolving power the band structure is revealed as closely spaced lines. As the transitions between the rotational levels are controlled by certain selection rules, the band structure is comparatively simple and not as complex as one might expect. The selection rules for the general case, when, for at least one of the two electronic states, the component of the electronic angular momentum along the line joining the two nuclei is *not* zero (i.e. $\Lambda \neq 0$, *vide infra*), is given by $\Delta J = 0$ or ± 1. For the case when for both the electronic states $\Lambda = 0$ (which is very common for the bands in the infra-red region) the selection rule is $\Delta J = \pm 1$, i.e. $\Delta J = 0$ is excluded. The series of lines for which $\Delta J = -1$ is called the P branch of the band, the series for which $\Delta J = +1$ is called the R branch and the series for which $\Delta J = 0$ the Q branch. In Fig. 8, the transitions corresponding to the P, R and Q branches are shown diagrammatically. It is to be noted that the

lines in the P and R branches form a single simple series of lines and can be represented by a single formula.

Term symbols.—Electrons in a diatomic molecule, unlike electrons in an atom, move in a cylindrically symmetrical, rather than in a spherically symmetrical field (the axis of symmetry lies along the line joining the nuclei). The compounded angular momentum vector L thus precesses

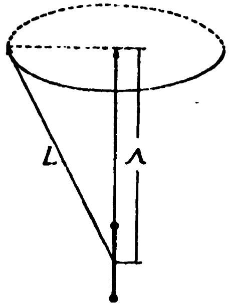


FIG. 9. In a diatomic molecule the two atomic nuclei provide a cylindrically symmetrical field with the symmetry axis along the line joining the nuclei. The orbital angular momentum L precesses about this line. The component of L along this line is Λ.

round this axis (Fig. 9). According to space quantization rule already discussed, the components of L along this axis are given by $M_L h/2\pi$ where M_L takes up values,

$$M_L = L, L-1, L-2, \ldots -L.$$

However, since in this case the energy is independent of signs, molecules with different $|M_L|$ only will have different energies. Further, on account of the strong interaction between the L's and the internuclear axis, states with different $|M_L|$ have greatly different energies. The electronic states of molecules are therefore appropriately classified according to their $|M_L|$ values. According to international nomenclature the symbol Λ is used for M_L. Λ can take up values,

$$\Lambda = 0, 1, 2, 3, \ldots L,$$

and the corresponding states are designated as

$$\Sigma, \Pi, \Delta, \Phi \ldots$$

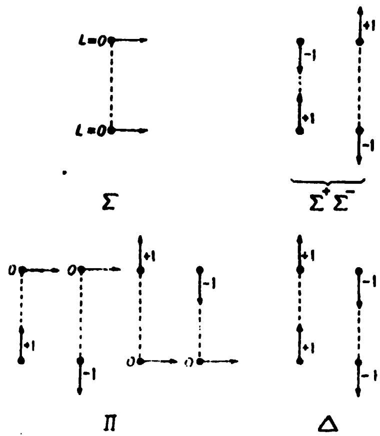


FIG. 10. Illustrating how the various molecular states may arise out of the combination of two 'separated' atoms. In the illustrations both the atoms are in the P state, i.e., $L_1 = L_2 = 1$. The components of L_1 and L_2 along the internuclear axis (dashed line) can thus have only the values -1, 0 and 1. The resultant L-values are 0, 1 and 2 and give rise to Σ, Π and Δ states.

In Fig. 10 a simple illustration is given of how the various molecular states may arise out of the combination of two 'separated' atoms. In the illustration, both the atoms are taken to be in the P-state, i.e. $L_1 = L_2 = 1$. The components M_{L_1} and M_{L_2} of L_1 and L_2 along the internuclear axis can thus have only the values -1, 0 and $+1$.

And, the resultant L has values, 0, 1 and 2 giving rise to Σ, Π and Δ states respectively.

Spin.—As in the case of atoms the spins combine to give a total spin momentum S with the quantum number S integral or half integral. The orbital motion of the electrons produces a magnetic field along the axis and S precesses round it. As before, the components of S along this axis are given by $M_s h/2\pi$. However, M_s for the molecule is designated Σ (not to be confused with Σ when $\Lambda = 0$) and we have for the values of Σ,

$$\Sigma = S, S-1, S-2 \ldots.$$

Note that for the state with $\Lambda = 0$, M_s cannot be defined, because there is no magnetic field.

Total angular momentum—multiplets.—As in the case of atoms the total angular momentum of the electron (designated Ω) is obtained by vector addition of Λ and Σ. However, since both Λ and Σ lie along the same line—the internuclear axis—simple algebraic addition is sufficient. We have

$$\Omega = |\Lambda + \Sigma|.$$

Thus Ω has $2S+1$ different values for a given value of Λ when $\Lambda > S$ and not equal to zero. Corresponding to these slightly differing values of Ω

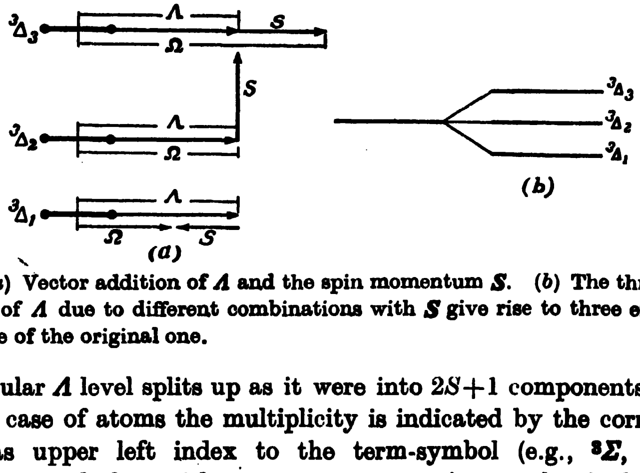


FIG. 11. (a) Vector addition of Λ and the spin momentum S. (b) The three modified values of Λ due to different combinations with S give rise to three energy levels in place of the original one.

the particular Λ level splits up as it were into $2S+1$ components (Fig. 11). As in the case of atoms the multiplicity is indicated by the corresponding number as upper left index to the term-symbol (e.g., $^3\Sigma$, $^2\Pi$, etc.). It may be noted that with one or two exceptions, only singlet ($S = 0$), doublet ($S = \frac{1}{2}$; $\Omega = \Lambda \pm \frac{1}{2}$) and triplet ($S = 1$; $\Omega = \Lambda, \Lambda \pm 1$) levels have been observed for diatomic molecules.

Complete term symbol.—The symmetry properties of the electronic eigenfunctions have also to be taken into account in the classification of the molecular electronic state. An electronic state is classified as + or as − according as the complete eigenfunction remains unchanged or changes in sign by reflection at the origin. This classification is indicated by superscripts (+) or (−) on the right of the term symbol.

An additional symmetry is possessed by homonuclear molecules, e.g., H_2, O_2, N_2, due to identical nuclear charges. The state is classified as *even*

(*gerade*) or *odd* (*ungerade*) according as the electronic eigenfunction remains unchanged or only changes in sign by reflection at the centre of symmetry (that is, by changing the co-ordinates of all the electrons, x_i, y_i, z_i to $-x_i$, $-y_i$, $-z_i$,). This classification is indicated by the subscripts g and u on the right of the term symbol. Thus, for example, we have a $^3\Sigma_u^+$ state, $^1\Delta_g^-$ state, etc. (*Note*: For states other than the Σ states, that is, for states for which $\Lambda \neq 0$, the (+) and the (−) terms have exactly the same value. Hence, it is usually unnecessary to make the (+, −) distinction.

For brevity the successive states of increasing energy values are often indicated arbitrarily by the letters, A, B, C . . . or a, b, c . . . , the ground state being indicated by the letter X. For ionized molecules the symbols are dashed, X', A', B'..., a', b'... etc. Thus the metastable state of the N_2 molecules at 6·1 eV (see Fig. 17) may be described as $A^3\Sigma_u^+$-state, $^3\Sigma_u^+$-state or simply as the A-state.

Intensity distribution in band spectra—the Franck-Condon Principle.— The intensities of the lines corresponding to the various sequences differ greatly from one another. The origin of these differences may be understood from the Franck-Condon Principle.

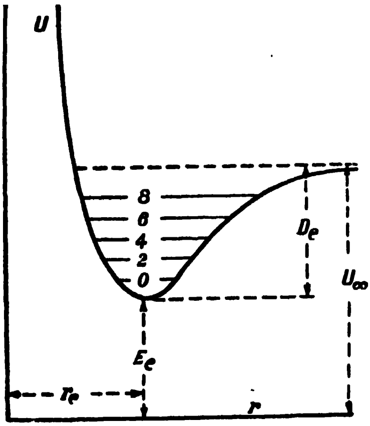


FIG. 12. Potential energy curve of a diatomic molecule. E_e is the electronic energy level. r_e is the mean distance between the nuclei. D_e is the energy above E_e for which the amplitude of vibration becomes infinite. Obviously, U_∞ is the dissociation energy $= E_e + D_e$.

Consider first the potential energy curve (U) due to nuclear separation and the vibrational states (v) of a diatomic molecule for a given electronic state E_e. In Fig. 12, r_e is the internuclear distance if there is no vibration. As the amplitude of the vibration increases the potential energy at the turning points increases, and for energy D_e above the energy level E_e, the amplitude becomes infinite. Obviously U_∞—the dissociation energy—is equal to $E_e + D_e$.

Now consider in Fig. 13 the two potential energy curves U' and U'' for the two energy states E_e' and E_e''. Take a v'-level (PP') belonging to the U' curve and drop a perpendicular from P to intersect the U'' curve at Q say. In this latter curve there is a v''-level (NN'') which intersects it at a point nearest to Q. Similarly, if we drop a perpendicular from P' to meet the U'' curve at Q' we can find a v''-level (MM'') which cuts the U'' curve at a

point nearest to Q'. *According to the Franck-Condon Principle transitions from the PP'-level will be more probable to these v''-levels (NN'' and MM'')*

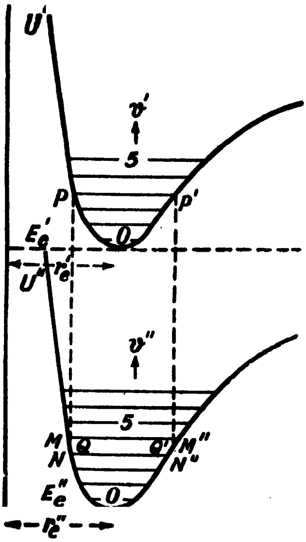


FIG. 13. Illustrating the Franck-Condon Principle. In the figure, perpendiculars dropped from the vibrational level PP' of the upper curve U' cut the lower curve U'' at QQ'. These points lie close to vibrational levels MM'' and NN''. According to the Franck-Condon Principle transitions from vibrational level PP' to MM'' and NN'' are more probable than to other levels.

than transitions to any other levels lying farther from Q or Q'. There are two reasons why these transitions are more probable than others: (*a*) An electronic transition takes place so rapidly that within this short interval the internuclear distance and the kinetic energy due to vibration do not change; (*b*) a particular electronic transition has a greater chance of occurring when the vibrating nuclei are in their extreme (maximum or minimum separations) positions because in course of vibration, the nuclei spend much more time at these two positions than at any other position. (It should be noted that the latter part of the second statement is strictly true when the levels are much above the zero level.) It is easily seen that the principle stated above, satisfies both these conditions.

(*c*) Metastable state

The time during which an atom or a molecule remains in an excited state is generally very small, of the order of 10^{-8}—10^{-9} sec. There are, however, certain excited states in which an atom or a molecule may remain for a much longer period—of the order of 10^{-2} sec. or more. Such excited states are known as metastable states. The metastable state arises from the fact that the spontaneous transition of the excited atom or molecule to a state of lower energy is 'forbidden' by selection rules. It is to be noted that by 'forbidden' it is not meant that the probability of such transition is zero but that it is very small. Two types of metastable states may be distinguished.

The electronic configuration of an atom or a molecule might change due to excitation, e.g., one of its outermost electrons might be lifted to

occupy an otherwise vacant orbit of higher energy. If transition from the excited orbit to an orbit of lower energy (generally the fundamental state) is forbidden by selection rules then the excited state is a metastable one.

Atoms having the outermost shell incomplete, may exist in different states for the same electronic configuration. For instance, oxygen atom has the electronic configuration $(1s^2)$ $(2s^2)$ $(2p^4)$. The outermost $2p$-shell holds four equivalent electrons (having the same n and the same l) though, according to Pauli's principle, it can hold six such electrons each with a different set of quantum numbers (see p. 604). These four electrons can be differently distributed amongst the six possible electronic levels and give rise to different energy states for the same electronic configuration $(1s)^2$ $(2s)^2$ $(2p)^4$ of the atom. Of these states, the one having the highest multiplicity is of the lowest energy and is the normal state. In the present example, the normal state is a ^3P one. There are also two possible states with higher energy, namely, ^1S and ^1D. Since transitions from these states to the normal state are forbidden, they are called metastable states of the oxygen atom. (See Fig. 17, Chapter X.)

It is thus seen that an atom in the metastable state is loaded with extra energy. Atoms with such extra energy and also with fairly long life can react easily with molecules or atoms of other type with which they come into contact. Transference of energy takes place and the colliding atoms or molecules may be ionized or excited.

(*d*) Radiationless decomposition—Predecomposition (Predissociation and Preionization)

The phenomenon of *radiationless decomposition* or *predecomposition* arises out of the overlapping of discrete energy levels belonging to one series with energy continuum belonging to another series as illustrated in

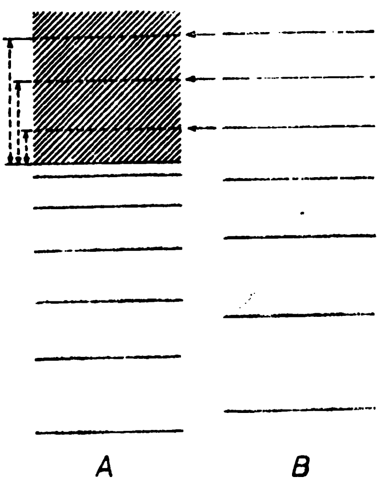


FIG. 14. Illustrating the phenomenon of radiationless decomposition or predecomposition arising out of the overlapping of discrete energy levels belonging to one series (B), with the energy continuum belonging to another series (A). Under such condition there are radiationless transitions from the discrete to the continuous states as indicated by horizontal arrows. The energy levels might be electronic, vibrational or rotational. For the first, the continuum represents ionization; for the second, it represents dissociation by vibration, and, for the third, dissociation by rotation.

Fig. 14. It will be noticed that in the series of levels in the higher energy state B, the level 5 and those above are overlapped by the continuum of

the lower energy state A. (The levels might be electronic or vibrational. For the former the continuum represents ionization; for the latter it represents dissociation.) According to quantum mechanical ideas if two different energy levels (belonging to two different series) have, by chance, approximately the same energy, they influence or perturb each other strongly. As a result of this perturbation, each of the two states assumes the properties of the other, that is, each is, as it were, made up of a mixture of the two states. (Another result of the perturbation is that the levels 'repel' each other; the higher level is pushed slightly upward and lower level slightly downward. But with this we are not immediately concerned here.) It is, therefore, clear from the figure that since in the continuum of A there are energy levels exactly equal to some of the levels of B (level 5 and above), there will be perturbation effect and the system, if it is in any of these discrete energy levels, will have a chance of going over to the continuum of A. There will thus be probability of decomposition (ionization or dissociation) by a radiationless process called *predecomposition* (*preionization* or *predissociation*).

It is to be noted that for the case of predissociation by vibration it is not just enough that the discrete vibrational levels of one electronic state are overlapped by the vibrational continuum of another. The two potential energy curves must cross each other at an appropriate point. It is further to be noted that the discrete levels in A may also be rotational levels (predissociation by rotation).

This predecomposition phenomenon is to be carefully distinguished from ordinary decomposition process in which (say, by absorption of light) there is direct transition to the continuous state. In predecomposition, on the other hand, there is first transition to a discrete energy state in which the system rests for a finite time after which it may either drop to a lower lying state by emission of light or may go over to the continuum of another state by which it is overlapped.

(e) Atomic nitrogen

The nitrogen atom has the electronic configuration $(1s)^2$ $(2s)^2$ $(2p)^3$. The three $2p$-electrons which form the incomplete L_2-cell are equivalent electrons, that is they have the same n and the same l ($= 1$). From the electron configuration given in Table I (row giving the m_l values) it is easily seen that the possible values of the resultant angular momentum of three of the electrons (out of the 6) are those corresponding to $L = 0$, 1 and 2, that is, the terms which the three electrons give rise to, are S, P and D respectively. Further, since the resultant spin (S) of the three can be either 3/2 or 1/2, the terms are either quartet or doublet (because multiplicity $= 2S+1$). Closer scrutiny shows that the S term is quartet and the P and D terms are doublets. Experiments, as also quantum-mechanical considerations, show that the term with the highest multiplicity has the lowest value. Hence ^4S is the lowest term of the N atom. ^2D and ^2P terms are of higher energy (2·37 and 3·56 eV respectively).

According to selection rules, inter-transitions between the terms ^4S, ^2D and ^2P are forbidden: hence ^2D and ^2P are metastable states (Fig. 15) It will be noticed that the ^2D state has the unusually long life of 8 hours!

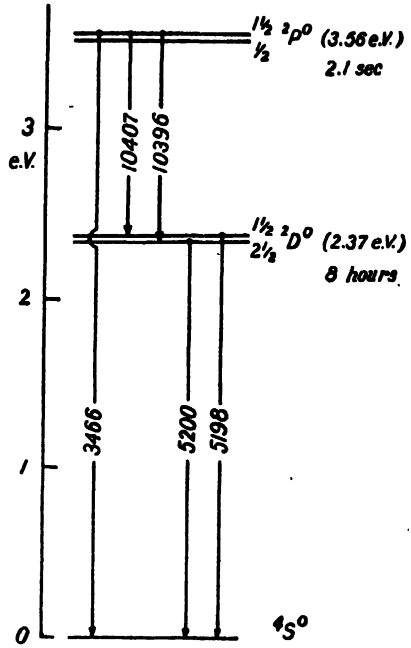


Fig. 15. Metastable energy levels of nitrogen atom. Note that the level $^2D^0$ has the extraordinarily long life of 8 hrs.! The lines λ 3466 and $\lambda\lambda$ 5198, 5200 have been observed in the aurora.

Forbidden transitions, however, do occur under certain special conditions as explained in Chapter X, Sec. 4 (*d*). In the auroral spectrum the forbidden transitions $^2P \rightarrow {}^4S$ (λ 3466) and $^2D \rightarrow {}^4S$ (λ 5200, λ 5198) have been identified. They have also been observed in Kaplan's specially conditioned discharge tubes. Their presence in the night air-glow spectrum is, however, doubtful.

It is to be noted that lines of atomic nitrogen due to permitted transitions $3p^4D \rightarrow 3s^4P$ (λ 8684 and λ 8714) have been observed in the aurora.

(*f*) Atomic oxygen

In the normal state the oxygen atom has the configuration $(1s)^2$ $(2s)^2$ $(2p)^4$. The four equivalent $2p$-electrons in the incomplete L_2-cell give rise to the terms ^3P, ^1D and ^1S (see above). Of these, the term ^3P with the highest multiplicity, has the lowest

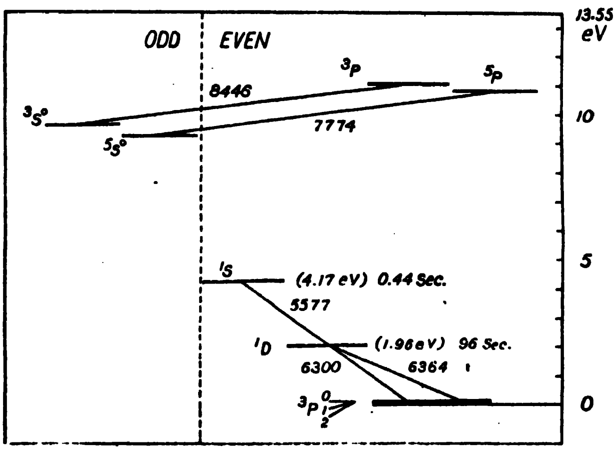


Fig. 16. Energy level diagram of O-atom, illustrating transitions as are of interest in night air-glow and auroral spectra. ^1S and ^1D levels are metastable. Transitions from the high energy levels ^3P and ^5p are allowed.

value and the corresponding state is the normal state of the atom. The energy-level diagram of oxygen atom depicting the states as are of interest in night air-glow and auroral spectra are depicted in Fig. 16. Selection rules demand that the inter-transitions between ^1S, ^1D and ^3P terms should be forbidden; but as already mentioned such forbidden transitions are known to occur under special conditions, e.g., in the upper atmosphere where the gas is in a very rarefied condition. The line λ 5577 ($^1S_0 \rightarrow {}^1D_2$) has been observed in the aurora, in the night air-glow spectrum and also in specially prepared discharge tube [Chapter X, Sec. 4(*d*)]. The lines λ 7774 and λ 8446 are due to permitted transitions and have been observed even when a faint trace of aurora is present.

Table II gives the absorption data of atomic oxygen. No experiment has yet been done to study its extreme ultraviolet absorption spectrum.

TABLE II

Absorption data of atomic oxygen

Spectral Region	Phenomena observed	Remarks
$\lambda < 910$ (13·55 eV)	O $^3P + h\nu \rightarrow O^{+\,4}S + e$ ..	The intensity of absorption is not known experimentally. Theoretical calculation shows that the absorption cross-section is of the order $2{\cdot}8 \times 10^{-17}$ cm.^{2}
$\lambda < 858$ (14·4 eV)	O $^1S + h\nu \rightarrow O^{+\,2}P + e$ ..	Effect negligible.
$\lambda < 744$ (16·5 eV)	O $^1D + h\nu \rightarrow O^{+\,2}P + e$ ..	Effect negligible.

(g) Molecular nitrogen

A nitrogen molecule may be made up of two atoms in any of the three states ^4S, ^2D, ^2P. This gives rise to a singlet and to a triplet series of terms (see Fig. 17). Not all the terms have, however, been observed. Those observed are,

Combination	Terms observed
$^4S + {}^4S$	$^1\Sigma$
$^2D + {}^2D$	$^3\Pi$
$^4S + {}^2D$	$^3\Sigma$
$^2D + {}^2P$	$^1\Pi$, $^3\Sigma$, $^3\Delta$, $^3\Pi$.

From Fig. 17 it will also be seen that the excited levels of the N_2 molecule fall under two categories—a set of singlet levels and a set of triplet levels. Intertransition between the two sets is forbidden by selection rules. Hence, the $A^3\Sigma$ state is a metastable state. The life of this

important metastable state has not yet been determined, though, it is believed to be long. (*Note*: In the energy-level diagram the singlet state

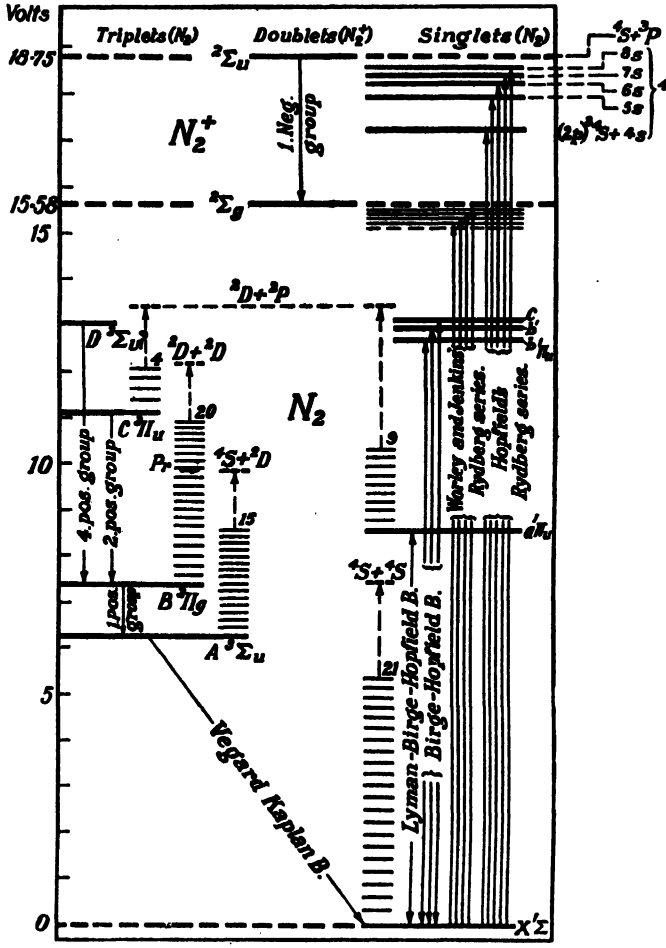


FIG. 17. Energy level diagram of N_2. The electronic states (a, b, c, A, B, C, D) are shown by the heavy horizontal lines. The short thin lines give the vibration levels. Two ionization limits 18·75 and 15·58 volts are shown by heavy horizontal broken lines. (*Note*: The singlet state $a^1\Pi$, as commonly believed, is shown as having u symmetry. According to Herzberg [2a], however, it most probably possesses a g symmetry.)

$a^1\Pi$, as commonly believed, is shown as having u symmetry. According to Herzberg [2a], however, it most probably possesses a g symmetry.)

The main singlet system of bands is the Lyman-Birge-Hopfield system. They lie below λ 2000 in the vacuum ultraviolet. Of the triplet system of bands, the strong ones, known as the first and the second positive systems, lie in the visible and near ultraviolet. Other systems of bands, less intense, are the fourth positive (triplet), the fifth positive (singlet) [2] and the Gaydon's

so-called *P*, *Q*, *R*, *S* and *T* systems (singlet) [2]. The transition of the fifth positive system is still uncertain (${}^1\Sigma_u^+ \rightarrow {}^1\Sigma_g^+$ or ${}^1\Sigma_g^- \rightarrow {}^1\Sigma_u^-$). The *P*, *Q*, etc. systems of bands are all due to transitions to $a^1\Pi_u$ state. These bands lie in the region λλ 2000–3500. (*Note*: The positive systems or bands are so called because they are prominent in the positive column of a discharge tube. It is interesting to note that there is no N_2 band system designated as third positive. This is because the system originally so designated was later found to be due to NO.)

There is an important intercombination system of bands (forbidden transition from triplet to singlet) known as the Vegard-Kaplan system lying in the visible and near ultraviolet.

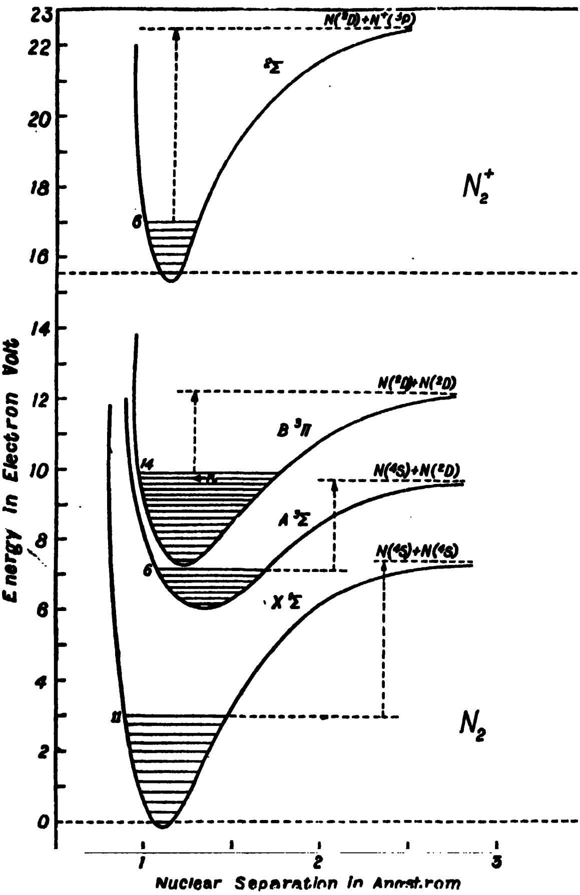


FIG. 18. Potential energy curves of the N_2 molecule.

Of the N_2^+ bands, the first negative bands lie in the visible and near ultraviolet. (The negative bands are so called because they are found strongest close to the cathode of a discharge tube.)

Of the absorption bands of N_2, the one of interest for excitation of upper atmospheric spectra, is the Hopfield-Rydberg series in the extreme

ultraviolet (λ 670) leading to ionization, the ionized molecule being in the excited ($A'\ {}^2\Sigma$) state [3]. In the absorption process one atom of the molecule remains unaffected and the other is excited as follows [4a]:

$$N^4S\,.\,N^4S(^1\Sigma) \rightarrow N^4S\,.\,N(1s)^2(2s)^2(2p)^2(ns)\ ^4P.$$

In the excited atom one of the three p-electrons passes to the ns states ($n = 3$, 4, 5, etc.) giving rise to a number of ^4P terms in a Rydberg series. The limit corresponds to the ionization of N_2 in the excited state $N^4S\,.N^{+3}P$ ($A'\ {}^2\Sigma$). [See Fig. 17.]

In Fig. 18 potential energy curves of the N_2 molecule in the different states are shown.

In Table III the absorption data of molecular nitrogen are summarized for convenience.

TABLE III

Absorption bands of N_2

Spectral Region	Remarks
$\lambda\lambda$ 1460–1000 ($^1\Pi \leftarrow {}^1\Sigma$) A progression of narrow discrete bands.	Weak absorption; about 40 cm. at S.T.P. is needed for complete absorption in the centre of the bands. The absorption is much weaker than that of O_2 in the same region. Observed by Birge and Hopfield [4].
$\lambda\lambda$ 1000–800 Rydberg sequence of absorption bands converging to the limit λ 795.	A few tenths of a mm. at S.T.P. cause complete absorption. First observed by Price [5]; also observed by Worley and Jenkins [6].
λ<795 Continuous absorption commencing at λ 795 (15·53 eV).	A few tenths of a mm. at S.T.P. show the continuum. Each quantum absorbed produces N_2^+ (normal) and electron. Observed by Worley and Jenkins [6].
$\lambda\lambda$ 800–660 Rydberg sequence of absorption bands converging to the limit λ 661.	A few hundredths of a mm. at S.T.P. suffice to bring out the absorption spectrum. First observed by Hopfield. Also observed by Takamine and others [7].
λ<661 Continuous absorption commencing at λ 661 (18·67 eV).	Very strong absorption as in the case of the preceding Rydberg sequence. Each quantum absorbed produces N_2^+ (excited) and electron. Observed by Takamine and others [7].

It will be noticed that molecular nitrogen, so far as the total amount in the upper atmospheric regions is concerned, is practically transparent from the visible region down to λ 1000.

Another important point may also be noticed. It is well known that in the case of most diatomic molecules (for instance in the case of O_2, Table V), there is a continuous absorption process which decomposes the molecule into a normal and an excited atom. There is, however, no such continuous

absorption of N_2 leading to the dissociation of the molecule into atoms. However, nitrogen molecules may be dissociated in the process of neutralization of N_2^+ ion by recombination with electron or, by predissociation (1150–1250 A). [See Chapter V, Sec. 3 and Chapter VIII, Sec. 6(*c*).]

In Table IV are listed the emission bands of N_2 and N_2^+ as are of importance in night air-glow and auroral spectra.

(*h*) Molecular oxygen

An oxygen molecule may be made up of two atoms in any of the three states ^3P, ^1S, ^1D. Fig. 19 shows the energy levels of O_2 and O_2^+. Corresponding to the four electronic states of O_2^+, namely, $X^2\Pi_g$, $a^4\Pi_u$, $a^2\Pi_u$, and $b^4\Sigma_g^-$, there are four different possibilities of direct photo-ionization, giving four—first, second, third and fourth—ionization potentials. (The ionization potentials are given in the Table at the end of Sec. 8.) It is to be noted that for all the cases, except for the ground state $X^2\Pi_g$, the absorption by the O_2 molecule, ionizes and, at the same time, raises the ionized molecule to an electronically excited state. (Compare the case of ionization of N_2, Fig. 17.)

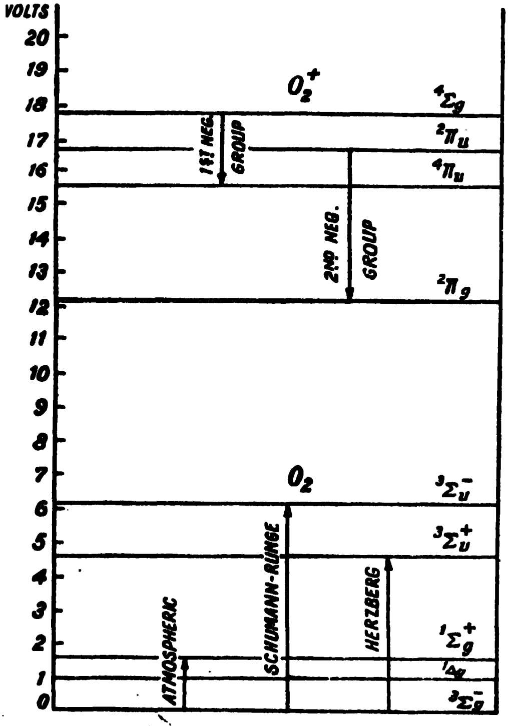


FIG. 19. Energy level diagram of normal and ionized oxygen molecule. (*Note*: According to Price and Collins [12] the distances of the $^4\Pi_u$ and $^4\Sigma_g^-$ states of O_2^+ from the ground state of O_2 are 16·1 and 18·2 eV respectively. See Table V.)

There is also a possibility of pre-ionization, attention to which was first directed by Nicolet [7(*a*)]. This pre-ionization process is important in the theory of the production of the *E*-region of the ionosphere (Chapter VI, Sec. 10). The Rydberg series (the limits of which lead to ionization) lying between 1000 A and 900 A in the bands for the second and third ionization processes, present characteristics of pre-ionization. These bands lie within the ionization continuum of the first ionization process. There is thus possibility of pre-ionization [*vide supra*, sub-sec. (*d*)]. The case is of importance, because the absorption within the bands is, at least, 10^5 times that of the continuum of the first ionization process within which they lie.

TABLE IV

Emission bands of N_2 *and* N_2^+

	Spectral region	Transition	Remarks
First positive (N_2)	λλ 14700—5030	$B^3\Pi_g \rightarrow A^3\Sigma_u$	Equally strong in both the night air-glow and the auroral spectra.
Second positive (N_2)	λλ 5440—2690	$C^3\Pi_u \rightarrow B^3\Pi_g$	Present in the auroral spectrum.
Fourth positive (N_2)	λλ 2904—2266	$D^3\Sigma_u \rightarrow B^3\Pi_g$	Not observed either in the auroral or in the night air-glow spectrum.
Vegard-Kaplan (N_2)	λλ 3400—2300	$A^3\Sigma_u \rightarrow X^1\Sigma_g$	Transition forbidden. Most intense in the night air-glow spectrum.
First negative (N_2^+)	λλ 5864—2987	$A'\,^2\Sigma \rightarrow X'\,^2\Sigma$	Prominent in auroral spectrum and also in that of the night air-glow when the upper atmosphere is illuminated by the solar rays.

The absorption coefficients for the ionization processes listed above are, unfortunately, not known with precision. The orders of the values as estimated by Nicolet [7(*b*)] are as follows:

First ionization	..	10^{-20} cm² ($\lambda < 1020$A)
Pre-ionization	..	10^{-16} to 10^{-15} cm² (1000A$<\lambda>$ 910A)
Second ionization	..	10^{-16} cm² ($\lambda < 770$)
Third ionization	..	10^{-17} cm² ($\lambda < 730$)
Fourth ionization	..	10^{-16} cm² ($\lambda < 680$)

As explained in Chapter VI, Sec. 10(*d*), the ionization processes of O_2, amongst those listed above, that are of interest in the formation of ionized regions are the first ionization ($\lambda < 1020$) and the bands of pre-ionization lying between 1000 A and 980 A.

The principal absorption bands that have been observed so far are given in Table V.

The emission bands which have so far been studied are: (1) the first negative bands λ 7900–5300 corresponding to the transition $^4\Sigma_g \rightarrow {}^4\Pi_u$ and (2) the second negative bands between λ 4400–2200 corresponding to the transition $^2\Pi_u \rightarrow {}^2\Pi_g$. It should be mentioned, however, that none of these bands have yet been observed in the night air-glow or auroral spectrum.

Potential energy curves for the O_2 molecule are shown in Fig. 20. The observed vibrational levels are indicated by the short horizontal lines.

TABLE V

Absorption bands of O_2

Spectral region: Bands observed	Remarks
λλ 7710—5380; $^1\Delta_g \leftarrow {}^3\Sigma_g^-$ $^1\Sigma_g^+ \leftarrow {}^3\Sigma_g^-$	Observed prominently in the solar spectrum due to atmospheric absorption; hence called atmospheric bands.
Fraunhofer *B*, *a*, *A* and *Z* bands.	Observed in night air-glow spectrum in absorption.
λλ 2600—2400 $^3\Sigma_u^+ \leftarrow {}^3\Sigma_g^-$ Herzberg bands [8].	Very feeble absorption due to forbidden transition. The bands appear to pass into a continuum at λ 2429 extending up to λ 2000 which may dissociate O_2 into two normal atoms in the ^3P state.
λλ 1925—1760 $^3\Sigma_u^- \leftarrow {}^3\Sigma_g^-$ Runge-Schumann bands [9].	2 to 3 cm. at S.T.P. produce strong absorption. According to Flory [10] the molecule is not only excited but may also be predissociated by absorption in the region. Collision of the excited O_2 with normal O_2 leads to the production of O_3. Warburg showed experimentally that the excited O_2 produced by absorption of one quantum of light $\lambda < 2040$ produces two O_3 molecules [Chapter IV, Sec. 4(*a*)].
λλ 1760—1350 Runge-Schumann continuum	Very strong absorption. 0·1 mm. at S.T.P. causes complete absorption. Leads to dissociation of O_2 into one ^3P (normal) and one ^1D (excited) atom [11] (see Chapter V).
λλ 1300—1019	Series of diffuse absorption bands tending to the limit λ 1019. A few tenths of a mm. at S.T.P. produce complete absorption. Observed by Price and Collins [12].
	Note.—A continuum beginning at the band limit (λ 1019) producing ionization (electron and normal O_2^+) is expected. The efficiency of this ionization appears to be low. The absorption in this region is very weak.
λλ 1000—670 Hopfield bands	Strongest of all the absorption bands of O_2. Observed by Hopfield and also by Price and Collins [12]. (i) Complicated system of absorption bands (*H*, *I*,. . .) which may be fitted into a Rydberg sequence converging to the limit λ 764 (16·14 eV). (ii) Another system of bands (*M*, *N*,. .) forming Rydberg sequence converging to the limit λ 744 (16·5 eV). An ionization continuum starting abruptly at about λ 740 (16·7 eV). (iii) A third system of bands (*P*, *Q*,. .) forming Rydberg sequence passing to the limit λ 670 (18·5 eV). Absorption in all these cases is very strong; a few hundredths of a mm. at S.T.P. is sufficient to show the bands clearly. The limits of all these electronic series correspond to ionization, the resulting ionized molecule being excited. It is also possible that there is pre-ionization of the molecule by absorption in the large wavelength end, λλ 910—1000.

(i) Negative ions

It is well known that free electrons in many gases attach themselves readily to the atoms or molecules of the gas. The atoms or molecules carrying the extra electronic charge are known as *negative ions*. If a negative ion is to be stable, the energy in the normal state must be less than that of the corresponding neutral atom. The difference of energy between the normal state of the neutral atom and of the ion is a measure of the 'electron affinity' of the atom. Investigations have shown the existence of O^-, O_2^-, NO_2^-, NO_3^-, OH^- and H^-. The ions N^-, N_2^-, He^-, Ne^- and A^- have not been observed [13]. The continuous spectrum of the negative ions lies in the visible and infra-red.

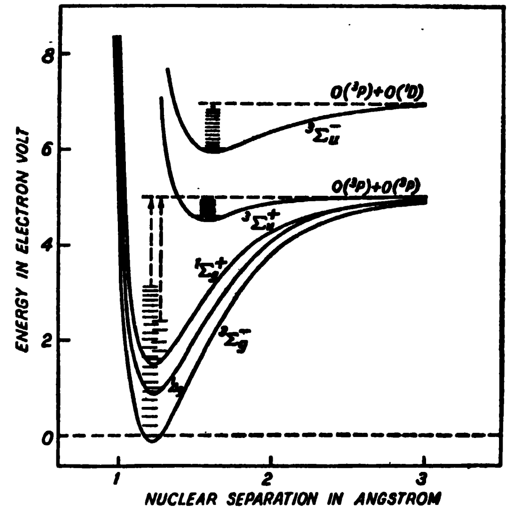


FIG. 20. Potential energy curves of the O_2 molecule. The observed vibrational levels are indicated by the short horizontal lines. Two dissociation energy levels are shown. In one the dissociation products are two $O(^3P)$ atoms and in the other (higher) one is $O(^3P)$ and the other $O(^1D)$. Absorption leading to the second type of dissociation is very much stronger.

Of the negative ions those of importance in the upper atmosphere are O^- and O_2^- [14].

The ground state of O^- has the electron configuration $(1s)^2(2s)^2(2p)^5$ and is a $^2P^0$ term. The ion may have also excited states but with very small binding energy. The value of the electron affinity of O^- is generally taken as 2·2 eV after Lozier [15]. Vier and Mayer's claim [16] that the value is 3·0 eV is not substantiated [16*a*]. The electron affinity of O_2^- is estimated to be 1 eV [14]. The absorption cross-section of O^- is of the order 10^{-8} cm.^{2}

3. CONDUCTIVITY AND DIELECTRIC CONSTANT OF AN ABSORBING IONIZED GAS

(a) Conductivity

The relation $\sigma = Ne^2\nu/m(p^2+\nu^2)$ in p. 183 can be proved as follows after Appleton and F. W. Chapman [*Proc. Phys. Soc.*, Vol. 44, p. 246, 1932].

The equation of motion of the electron under the action of a sinusoidally varying electric force can be written as

$$m\frac{d^2x}{dt^2} = E_0 e\epsilon^{jpt} \quad \dots \quad \dots \quad \dots \quad \dots \quad (1)$$

where the symbols have their usual meanings.

Integrating,

$$\frac{dx}{dt} = v = \frac{-jE_0 e \epsilon^{jpt}}{mp} + C \quad \ldots \quad \ldots \quad \ldots \quad (2)$$

where C is an integration constant.

Consider the group of electrons which had all their last collision at a particular instant t_1. Since the mean velocity of the group of electrons at the instant t_1 will be zero,

$$\left(\frac{dx}{dt}\right)_{t=t_1} = 0.$$

Then from Eq. (2)

$$C = \frac{jE_0 e}{mp} \epsilon^{jpt_1}. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (3)$$

Each of the electrons of this group can, therefore, be assigned a velocity at the time t, given by,

$$v = \frac{-jE_0 e}{mp}\left[\epsilon^{jpt} - \epsilon^{jpt_1}\right]$$

$$= \frac{-jE_0 e}{mp}\left[1 - \epsilon^{-jp(t-t_1)}\right]\epsilon^{jpt}. \quad \ldots \quad \ldots \quad (4)$$

We now find out the mean velocity of *all* the electrons, the regular vibrations of which are being interrupted by collisions. For this, we first find out the number of particles whose mean velocity lies in the neighbourhood of v, that is the number of particles whose time interval between last collision and the time t, lies in the range $(t-t_1)$ and $(t-t_1)+dt_1$. If the average time interval between two collisions be τ, then according to Lorentz the number would be

$$\frac{N_0}{\tau}\epsilon^{\frac{-(t-t_1)}{\tau}} dt_1$$

where N_0 is the total number of particles considered. Since t_1 varies from $t_1 = -\infty$, to $t_1 = t$, we have for the mean velocity of all the particles, the expression,

$$\bar{v} = \frac{1}{N_0}\int_{-\infty}^{t} \frac{-jE_0 e}{mp}\left[1 - \epsilon^{-jp(t-t_1)}\right]\epsilon^{jpt}\frac{N_0}{\tau}\epsilon^{\frac{-(t-t_1)}{\tau}} dt_1$$

$$= \frac{-jE_0 e}{mp\tau}\epsilon^{jpt}\int_{-\infty}^{t}\left[\epsilon^{\frac{-(t-t_1)}{\tau}} - \epsilon^{-(t-t_1)\left(jp+\frac{1}{\tau}\right)}\right] dt_1$$

$$= \frac{E_0 e \epsilon^{jpt}}{jmp\tau}\cdot\left(\tau - \frac{1}{jp+\frac{1}{\tau}}\right)$$

$$= \frac{E_0 e \epsilon^{jpt}}{m\left(jp+\frac{1}{\tau}\right)} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (5)$$

Putting $1/\tau = \nu$, the collisional frequency,

$$\bar{v} = \frac{E_0 e\epsilon^{jpt}}{m(p^2+\nu^2)}(\nu - jp). \quad \ldots \quad \ldots \quad \ldots \quad (6)$$

Assuming that there are N electrons per cm³. the current density is $Ne\bar{v}$. This current being due to the electric field $E_0\epsilon^{jpt}$, the conductivity of the medium is obtained by dividing $Ne\bar{v}$ by $E_0\epsilon^{jpt}$. Retaining the real part of the complex expression, the conductivity

$$\sigma = \frac{Ne^2\nu}{m(p^2+\nu^2)} \text{ [e.s.u.]}. \quad \ldots \quad \ldots \quad \ldots \quad (7)$$

The effect of the collisions in disturbing the regular vibrations may also be conveniently represented as a frictional effect. Thus, if a frictional coefficient g be assumed for the vibrating electron, the equation of motion becomes

$$m\frac{d^2x}{dt^2}+g\frac{dx}{dt} = E_0 e\epsilon^{jpt} \quad \ldots \quad \ldots \quad \ldots \quad (8)$$

or,

$$m\frac{dv}{dt}+gv = E_0 e\epsilon^{jpt}. \quad \ldots \quad \ldots \quad \ldots \quad (9)$$

Since v varies sinusoidally with the same frequency as the impressed electric force, we may put $dv/dt = jpv$. The expression for the velocity thus becomes:

$$\bar{v} = \frac{E_0 e\epsilon^{jpt}}{jpm+g}. \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (10)$$

Comparing Eq. (10) with Eq. (5) we note that the frictional coefficient

$$g = \frac{m}{\tau} = m\nu.$$

(b) Complex dielectric constant

The expression for the complex dielectric constant of an absorbing medium and the relation between the dielectric constant and the refractive index of such a medium [Chapter VI, Sec. 2] can be derived as follows:—

Consider the propagation of a plane electromagnetic wave in a medium of dielectric constant κ and conductivity σ in the direction of the X-axis (say). Let E_x, E_y, E_z denote the electric vectors due to the passing wave along the X, Y and Z axes. The total current along the Y-axis due to the wave is composed of two parts, a displacement current $\kappa/4\pi.\partial E_y/\partial t$ due to the time variation of E_y and a conduction current $E_y\sigma$. Let E_y be of the form $E_{0y}\epsilon^{jpt}$ Then the total current may be expressed as

$$\frac{\kappa}{4\pi}\frac{\partial E_y}{\partial t}+\frac{\sigma}{jp}\frac{\partial E_y}{\partial t} = \frac{\kappa - \dfrac{4\pi j\sigma}{p}}{4\pi}\frac{\partial E_y}{\partial t}.$$

The conductivity may thus be regarded to have altered the dielectric constant of the medium from κ to a complex value

$$\kappa' = \kappa - \frac{4\pi j\sigma}{p}. \quad \ldots \quad (1)$$

The equation for the propagation of E_y may be now written as,

$$E_y = E_{0y} e^{jp\left(t - \frac{\sqrt{\kappa'}}{c}x\right)}. \quad \ldots \quad (2)$$

where c is the velocity of light.
Or, putting the complex quantity

$$\sqrt{\kappa'} = \mu - j\chi, \quad \ldots \quad (3)$$

$$E_y = E_{0y} e^{-\frac{p\chi}{c}x} \cdot e^{jp\left(t - \frac{\mu}{c}x\right)}.$$

μ the real part of $\sqrt{\kappa'}$ is thus the refractive index of the medium, while χ the imaginary part is the absorption index, causing a diminution of the amplitude as the wave progresses. χ is the absorption per length of path $c/p = \lambda/2\pi$. Or, if k is the absorption per unit length,

$$\chi = ck/p. \quad \ldots \quad (4)$$

Now in an ionized medium there is, a *convection* current, due to the motion of the ions under the action of the electric field of the wave, the components of which are $Nev_x = d/dt(Nex)$, etc. The convection current, it is to be noted, is *in* phase with the displacement current and does not cause any dissipation of energy. In fact, as mentioned in p. 179, the effect of the convection current, so far as the propagation of a wave of pulsatance p is concerned, is merely to reduce the dielectric constant of the space to the value $1-(4\pi Ne^2/mp^2)$. Remembering, therefore, that for plane wave propagation in the direction of the X-axis, $\partial/\partial y = \partial/\partial z = 0$, the field equations may be written in the form

$$\frac{1}{c}\left[\frac{\partial E_x}{\partial t} + \frac{\partial}{\partial t}(4\pi Nex)\right] = 0 \quad \ldots \quad (5a)$$

$$\frac{1}{c}\left[\frac{\partial E_y}{\partial t} + \frac{\partial}{\partial t}(4\pi Ney)\right] = -\frac{\partial H_z}{\partial x} \quad \ldots \quad (5b)$$

$$\frac{1}{c}\left[\frac{\partial E_z}{\partial t} + \frac{\partial}{\partial t}(4\pi Nez)\right] = \frac{\partial H_y}{\partial x} \quad \ldots \quad (5c)$$

$$-\frac{\partial H_x}{\partial t} = 0, \; -\frac{1}{c}\frac{\partial H_y}{\partial t} = -\frac{\partial E_z}{\partial x}, \; -\frac{1}{c}\frac{\partial H_z}{\partial t} = \frac{\partial E_y}{\partial x}. \quad \ldots \quad (5d)$$

We note from Eq. (5*a*) the curious fact that the wave is not wholly transverse since $E_x \neq 0$, though the wave propagation has been assumed to be in the direction of the X-axis.

Let the effective dielectric constant of the medium be κ. If there be absorption due to collision, then κ will be complex and may be represented

by κ'. Let the vectors E and H vary sinusoidally. Then, according to Eq. (2) E_z, for instance, will contain the factor

$$\epsilon^{jp\left(t-\frac{\sqrt{\kappa'}}{c}x\right)} \text{ and we have } \frac{\partial}{\partial x} = -\frac{jp\sqrt{\kappa'}}{c} \text{ and } \frac{\partial}{\partial t} = jp.$$

Substituting these values in the set of Eqs. (5) we have,

$$jpE_x + 4\pi jpNex = 0$$

$$\frac{1}{c}\left[jpE_y + 4\pi jpNey\right] = \frac{jp\sqrt{\kappa'}}{c} H_z \qquad .. \qquad .. \qquad (6)$$

$$\frac{1}{c}\left[jpE_z + 4\pi jpNez\right] = \frac{-jp\sqrt{\kappa'}}{c} H_y$$

$$-jpH_x = 0, \; -\frac{1}{c} jpH_y = \frac{jp\sqrt{\kappa'}}{c} E_z, \; -\frac{1}{c} jpH_z = -jp\frac{\sqrt{\kappa'}}{c} E_y \qquad .. \qquad (7)$$

Eliminating H_z from Eq. (6) with the help of Eq. (7) we have the relation:

$$E_y(\kappa'-1) = 4\pi Ney \qquad .. \qquad .. \qquad .. \qquad .. \qquad (8a)$$

Similarly,

$$E_z(\kappa'-1) = 4\pi Nez \qquad .. \qquad .. \qquad .. \qquad .. \qquad (8b)$$

$$H_y(\kappa'-1) = -\sqrt{\kappa'}4\pi Nez \qquad .. \qquad .. \qquad .. \qquad (9a)$$

$$H_z(\kappa'-1) = \sqrt{\kappa'}4\pi Ney \qquad .. \qquad .. \qquad .. \qquad (9b)$$

(These expressions will be found helpful in connection with the deduction of the Appleton-Hartree formula in Sec. 4.)

Now, the equation of motion of the ion under influence of the electric vector E_y (say), is given by

$$m\frac{\partial^2 y}{\partial t^2} = eE_y - g\frac{\partial y}{\partial t}. \qquad .. \qquad .. \qquad .. \qquad (10)$$

There is also a similar equation for E_z. The second term on the right represents the effect of 'friction' on the motion of the ion due to collisions where $g = m\nu$.

[A force arising out of polarization is sometimes included. The total electric force acting on the ion along Y-axis (for instance), then becomes,

$$E_y + \frac{4\pi}{3} Ney.$$

It has, however, been shown by Darwin that under ionospheric conditions, this force does not exist. The subject is discussed in Chapter VI, Sec. 4(f). There may also be a force of restitution due to the displacement of the ion. This has been neglected in writing out Eq. (10), as its magnitude must be very small, since the ion is moving freely.]

Substituting for $\partial^2/\partial t^2$ and $\partial/\partial t$ we have, $-mp^2y = eE_y - gjpy$. Or, substituting for E_y from Eq. (8a), we have,

$$\frac{1}{\kappa'-1} = -\frac{mp^2}{4\pi Ne^2} + \frac{jgp}{4\pi Ne^2} = \alpha + j\beta,$$

where
$$\alpha = -\frac{mp^2}{4\pi Ne^2} \text{ and } \beta = \frac{pg}{4\pi Ne^2} = \frac{mp\nu}{4\pi Ne^2}.$$

Thus, when the effect of absorption due to collision is taken into account, the value of effective (complex) dielectric constant κ' of the ionized medium [see Eqs. (3) and (4)] is given by

$$\kappa' = M^2 = \left(\mu - \frac{jck}{p}\right)^2 = 1 + \frac{1}{\alpha + j\beta}$$

where M is the complex refractive index.

Since according to Eq. (1), $\kappa' = \kappa - 4\pi j\sigma/p = (\mu - j\chi)^2$ it is easy to show by equating the real and the imaginary parts that the refractive index μ and the dielectric constant κ are related by the equation

$$\mu^2 = \frac{\kappa}{2} + \sqrt{\frac{\kappa^2}{4} + \frac{4\pi^2\sigma^2}{p^2}}.$$

When the conductivity σ is equal to zero we can identify μ^2 with κ.

4. THE APPLETON-HARTREE MAGNETO-IONIC FORMULA

The Appleton-Hartree formula [Chapter VI, Sec. 2] may be derived as follows:—

Let the propagation be along the X-axis and let the Z-axis be in the direction of H_T (Fig. 14, p. 195). The equations of motion of the electron in the field of the advancing electromagnetic wave are then:

$$m\frac{d^2x}{dt^2} = eE_x - g\frac{dx}{dt} + \frac{e}{c}H_T\frac{dy}{dt}. \quad \ldots \quad \ldots \quad \ldots \quad (1)$$

$$m\frac{d^2y}{dt^2} = eE_y - g\frac{dy}{dt} - \frac{e}{c}H_T\frac{dx}{dt} + \frac{e}{c}H_L\frac{dz}{dt} \quad \ldots \quad \ldots \quad (2)$$

$$m\frac{d^2z}{dt^2} = eE_z - g\frac{dz}{dt} - \frac{e}{c}H_L\frac{dy}{dt}. \quad \ldots \quad \ldots \quad \ldots \quad (3)$$

Here g is the 'frictional' coefficient [see Sec. 3(*a*)].
Making the substitutions, $d^2/dt^2 = -p^2$, $d/dt = jp$

we have, $$-mp^2x = eE_x - gjpx + \frac{e}{c}H_T\,jpy \quad \ldots \quad \ldots \quad (4)$$

and similar equations for y and z.

$$\therefore E_x = -\frac{mp^2x}{e} + \frac{gjpx}{e} - \frac{H_T}{c}jpy \quad \ldots \quad \ldots \quad \ldots \quad (5)$$

or, substituting for α, β and γ_T, remembering that g is equal to $m\nu$, we have

$$E_x = 4\pi Nex(\alpha + j\beta) - j\gamma_T(4\pi Ney). \quad \ldots \quad \ldots \quad \ldots \quad (6)$$

Similarly,

$$E_y = 4\pi Ney(\alpha + j\beta) + j\gamma_T(4\pi Nex) - j\gamma_L(4\pi Nez) \quad \ldots \quad (7)$$

and,

$$E_z = 4\pi Nez(\alpha + j\beta) + j\gamma_L(4\pi Ney). \quad \ldots \quad \ldots \quad \ldots \quad (8)$$

We also have from Eqs. (8) in Sec. 3,

$$\left.\begin{aligned} -E_x &= 4\pi Nex \\ (\kappa'-1)E_y &= 4\pi Ney \\ (\kappa'-1)E_z &= 4\pi Nez \end{aligned}\right\} \quad \ldots \quad (9)$$

Substituting the values of E_x, E_y and E_z from Eqs. (6), (7) and (8) to (9) we have,

$$4\pi Nex = \frac{j\gamma_T 4\pi Ney}{(1+\alpha+j\beta)} \quad \ldots \quad (10)$$

$$Ney\left[\frac{1}{\kappa'-1}-(\alpha+j\beta)+\frac{\gamma_T^2}{1+\alpha+j\beta}\right]=j\gamma_L Nez. \quad \ldots \quad (11)$$

$$Nez\left[\frac{1}{\kappa'-1}-(\alpha+j\beta)\right]= -j\gamma_L Ney. \quad \ldots \quad (12)$$

Multiplying Eqs. (11) and (12), we have,

$$\gamma_L^2 = \left[\frac{1}{\kappa'-1}-(\alpha+j\beta)\right]\left[\frac{1}{\kappa'-1}-(\alpha+j\beta)+\frac{\gamma_T^2}{(1+\alpha+j\beta)}\right]. \quad \ldots \quad (13)$$

This equation may be thrown into the standard form $ax^2+bx+c=0$, and written as,

$$\left[\frac{1}{\kappa'-1}-(\alpha+j\beta)\right]^2+\frac{\gamma_T^2}{1+\alpha+j\beta}\left[\frac{1}{\kappa'-1}-(\alpha+j\beta)\right]-\gamma_L^2=0. \quad \ldots \quad (14)$$

Solving Eq. (14) we obtain,

$$\kappa' = M^2 = 1+\frac{1}{(\alpha+j\beta)-\dfrac{\gamma_T^2}{2(1+\alpha+j\beta)}\pm\sqrt{\dfrac{\gamma_T^4}{4(1+\alpha+j\beta)^2}+\gamma_L^2}}. \quad \ldots \quad (15)$$

5. A NOTE ON THE DIFFERENT WAYS OF EXPRESSING THE PROBABILITY OF A REACTION

The probability of a reaction may be expressed in various ways [Chapter VI, Sec. 12]. For instance, one may say of *probability of the reaction taking place per gas kinetic collision*, meaning thereby, the probability of the reaction taking place if the reacting particles collide in the gas kinetic sense. The same fact might also be expressed as the *effective cross-section* of the particle for the particular reaction. The value of the cross-section is such that if the frequency of collision be calculated with the help of the gas kinetic formula (taking this value of the cross-section) then every such collision would be effective in producing the reaction. Evidently the probability of the reaction per gas kinetic collision is equal to the ratio of the 'effective' cross-section and the gas kinetic cross-section. Finally, we may express the efficiency by the *coefficient* of the process (recombination, mutual neutralization, attachment, detachment, etc.). Thus, if n_1 and n_2 be the numbers per unit volume of the two kinds of particles 1 and 2, then the

rate of the reaction, i.e., the number of the reactions per unit time per unit volume will be given by $\alpha n_1 n_2$ where α is called the coefficient of the reaction.

The relations between the 'effective' cross-section of the particle, the 'coefficient' of the reaction and the 'probability of the reaction taking place per gas kinetic collision' may be deduced as follows:

Let ν_1 be the number of collisions which a particle of kind n_1 make per second. ν_1 is given by $\bar{v}/l_1$ where $\bar{v}$ is the mean velocity and l_1 the mean free path of particles of kind n_1. l_1 is given by

$$l_1 = \frac{1}{\pi n_2 r_{12}^2 \left(1 + \frac{\overline{C_2^2}}{\overline{C_1^2}}\right)^{\frac{1}{2}}}$$

where $r_{12} = r_1 + r_2$ the sum of the radii of the two kinds of particles and $\overline{C_1^2}$ and $\overline{C_2^2}$ their respective mean square velocities. If the particles are of the same kind, that is, if the collisions are between the molecules of a single gas, then $\overline{C_1^2} = \overline{C_2^2}$ and $r_{12} = 2r$. In such case

$$l = \frac{1}{4\pi n r^2 \sqrt{2}} = \frac{1}{4\sqrt{2}\sigma n},$$

where σ is the cross-section of the particle. If one kind of the particles be electrons, then their mean free path

$$l_e = \frac{1}{n\sigma}.$$

This is because the radius of the electron is negligible compared to that of the other kind of particles and, for the temperatures considered, the mean square velocity of the electron $\overline{C_1^2}$ is much larger than that of the other velocity $\overline{C_2^2}$, so that $\overline{C_2^2}/\overline{C_1^2} < 1$. Therefore, since

$$\bar{v} = \sqrt{\frac{8kT}{\pi m}},$$

the collisional frequency, when the particles are of the same kind, is given by

$$\nu = \frac{\bar{v}}{l} = 16\sqrt{\frac{kT}{\pi m}}\,\sigma n,$$

and, if one kind of the particles be electrons, then

$$\nu_e = \sqrt{\frac{8kT}{\pi m}}\,n\sigma.$$

The frequency of collision as deduced above is that of the gas kinetic collision and the cross-sections of the colliding particles are the gas kinetic cross-sections. The latter is of the order of 3×10^{-16} cm.^{2}

and varies according to the nature of the particle. It can be determined with accuracy by viscosity experiments.

The cross-section may also be the *effective* cross-section of a particle for a particular kind of reaction so that,

$$\left.\begin{matrix}\text{the probability of the reaction taking}\\ \text{place per gas kinetic collision}\end{matrix}\right\} = \frac{\text{effective cross-section}}{\text{gas kinetic cross-section}}.$$

The collisional frequency calculated by taking the effective cross-section may be called the effective collisional frequency of the particle for the reaction considered.

We may also find out the relations between the 'coefficient' of the process α, the 'effective' collisional frequency $\nu_{\text{eff.}}$ and the gas kinetic cross-section σ.

We have,

$$\left.\begin{matrix}\text{number of reactions per sec.}\\ \text{per unit volume}\end{matrix}\right\} = \alpha n_1 n_2.$$

Also,

$$\left.\begin{matrix}\text{number of effective collisions per sec. of particles}\\ \text{of kind } n_1 \text{ in unit volume leading to the reaction}\end{matrix}\right\} = \nu_{\text{eff.}}.n_1.$$

Since each of these collisions is successful in producing the reaction,

$$\nu_{\text{eff.}}.n_1 = \alpha n_1 n_2,$$

or,

$$\alpha = \frac{\nu_{\text{eff.}}}{n}.$$

Since,

$$\nu_{\text{eff.}} = 16\sqrt{\frac{kT}{\pi m}}\, n\sigma_{\text{eff.}} \text{ when the particles are of the same kind,}$$

$$= \sqrt{\frac{8kT}{\pi m}}\, n\sigma_{\text{eff.}} \text{ when one kind of the particles is electron,}$$

the coefficient of the reaction is given by,

$$\alpha = 16\sqrt{\frac{kT}{\pi m}}\, \sigma_{\text{eff.}} \text{ when the particles are of the same kind,}$$

$$= \sqrt{\frac{8kT}{\pi m}}\, \sigma_{\text{eff.}} \text{ when one kind of the particles is electron.}$$

As an illustration we may calculate the attachment probability and the effective cross-section of O atom for the reaction:

$$O + e \rightarrow O^- + h\nu$$

from the knowledge of its coefficient.

Theoretical considerations show that the coefficient of the reaction (β) is 2×10^{-15} cm.^{3}/sec. for electrons of energy 0·1 eV. The effective cross-section is given by

$$\beta = \bar{v}\,.\,\sigma_{\text{eff.}}$$

and since $\frac{1}{2}mv^2 = eV$ where V is the energy of the colliding electron,

$$\beta = \sqrt{\frac{2eV}{m_e}}\,\sigma_{eff.}$$

$$\text{or, } \sigma_{eff.} = \sqrt{\frac{m_e}{2eV}}\,\beta = 4{\cdot}3\times10^{-24}\ \text{cm.}^2$$

Again, the probability of the reaction taking place per gas kinetic collision is given by,

$$\frac{\text{effective cross-section}}{\text{gas kinetic cross-section}} = \frac{4{\cdot}3\times10^{-24}}{3\times10^{-16}} \approx 1{\cdot}4\times10^{-8}.$$

If ν the number of gas kinetic collisions taking place per sec. be known, then the number of the reactions taking place per second is obviously $\nu\times1{\cdot}4\times10^{-8}$. It is important to remember that ν here refers to gas kinetic collisions between electron and oxygen atoms. If there be particles of other kind present with which the electron can also collide, e.g., N_2 molecules, and if ν refers to the total number of gas kinetic collisions suffered by the electron, then in order to obtain the number of reactions per second, ν has to be multiplied by the factor n_1/n where n_1 is the number per unit volume of oxygen atoms and n that of all kinds of particles. Thus, it is known from radio measurements that the frequency of collision of electrons in the lower portion of Region E is about 10^6 per second. Of these, if one-fifth be taken to be with O atoms (the other kind of particles being O_2 and N_2), and since for Region E gas kinetic cross-section is roughly equal to elastic cross-section, the rate of the reaction is given by $\frac{1}{5}\times10^6\times1{\cdot}4\times10^{-8} = 2{\cdot}8\times10^{-3}$ per second [Chapter VI, Sec. 15].

6. RATE OF PHOTO-DETACHMENT OF ELECTRONS FROM NEGATIVE IONS BY SOLAR RADIATION

In Chapter VI, Sec. 12, it is mentioned that I, the rate of photo-detachment of electrons from negative ions under the influence of solar radiation is given by $I = N_-\gamma_1\,\beta$ where N_- is the number density of the negative ions and β is the coefficient of attachment. The expression for γ_1 as given in the text may be deduced as follows.

The number of quanta of energy $h\nu$ in the frequency range $\nu\rightarrow\nu+d\nu$ entering into unit area of earth's atmosphere is given by,

$$n_\nu d\nu = \frac{2\pi}{c^2}\left(\frac{R}{d}\right)^2 \frac{\nu^2}{e^{h\nu/kT_s}-1}\,d\nu$$

where

R—radius of the sun,

d—earth-sun distance,

T_s—temperature of the sun.

Since for the conditions under consideration $1 \ll \exp h\nu/kT_s$ the expression reduces to

$$n_\nu d\nu = \psi^2 \frac{2\pi\nu^2}{c^2} \exp\left(-h\nu/kT_s\right)d\nu,$$

where ψ is the angle subtended by the sun's radius at the earth.

Now the rate of photo-detachment, i.e., the number of electrons detached from negative ions by the impact of radiation per unit volume per second is given by,

$$I = N_- \int_{\nu_0}^{\infty} Q_a(\nu)n_\nu d\nu,$$

where,

$Q_a(\nu)$—cross-section for light absorption of frequency ν.

ν_0—light frequency for the minimum energy of detachment.

N_-—number of negative ions per cm^3.

Substituting in the expression of I the value of n_ν and using Milne's relation (p. 303) we have,

$$I = \frac{\pi m^2 N_- \psi^2}{h^2} \int_{\nu_0}^{\infty} Q_e(\nu)v^2 \exp\left(-h\nu/kT_s\right)d\nu.$$

Now changing the variable from ν to v with the help of the relation,

$$\tfrac{1}{2}mv^2 = h(\nu-\nu_0),$$

and remembering that when ν is small, $Q_e(v) \propto 1/v$ [1], i.e., $Q_e(v) = Q'_{e0}/v$, where Q_{e0} is a constant, we have,

$$I = \frac{\pi m^3 N_- \psi^2 Q_{e0}}{h^3} \exp\left(-h\nu_0/kT_s\right) \int_0^{\infty} v^2 \exp\left(-\frac{mv^2}{2kT_s}\right)dv.$$

Now the product of $Q_e(v)$ and v is the coefficient of attachment β for electrons with velocity v. We therefore have from the relation $Q_e(v) = Q_{e0}/v$ that $\beta = Q_{e0}$. Substituting β for Q_{e0} in the expression of I and integrating

$$I = N_- \psi^2 \frac{(2\pi m)^{\frac{3}{2}}}{4h^3} k^{\frac{3}{2}} T_s^{\frac{3}{2}} \beta \exp\left(-h\nu_0/kT_s\right).$$

Putting $I = N_- \gamma_1 \beta$ we have,

$$\gamma_1 = \psi^2 \frac{(2\pi m)^{\frac{3}{2}}}{4h^3} k^{\frac{3}{2}} T_s^{\frac{3}{2}} \exp\left(-h\nu_0/kT_s\right).$$

7. A NOTE ON GEODETIC CURVES

Some of the properties of geodetic curves mentioned in Chapter VIII, Sec. 5(c), are briefly discussed here.

A geodetic curve on a cone of revolution is a curve such that when the cone is rolled on a plane the trace of the curve on the plane is a straight line [1]. Or, conversely, if a straight line is drawn on the plane developed

by rolling the cone, then when the plane is wound up into the form of the original cone, the line is a geodetic line. In Fig. 1 the radial lines mark out the spaces over which the cone advances with each revolution. The angle ϕ between the lines is easily seen to be equal to $2\pi \sin \omega$ where ω is the semi-apical angle. The line AB on the plane will be a geodetic line when the plane is wound up to re-form the cone. A little consideration shows that the point where the line is at a minimum distance d from the apex will be the turning point of the geodetic line on the surface of the cone. That is, the line will come spiralling down the surface of the cone up to this point and will then spiral up again.

Consider a particle of characteristics m, v, e moving in a uniform magnetic field (H) at right angles to the lines of force. Due, as usual, to the balancing of the electromagnetic force and the centrifugal force the particle will move in a circular path of radius $r_0 = mv/eH$, the plane of the circle being at right angles to the lines of force.

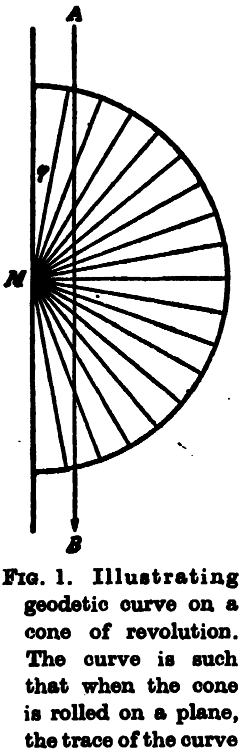


FIG. 1. Illustrating geodetic curve on a cone of revolution. The curve is such that when the cone is rolled on a plane, the trace of the curve on the plane is a straight line (AB).

If the particle be not moving in a direction at right angles to the lines of force but in a direction making an angle ψ with the same then the velocity component parallel to the magnetic field will remain unaltered and the particle will move along a helix, with radius of curvature ρ, where $\rho = mv/eH \sin \psi$, the axis of the helix being parallel to the magnetic field. This helix would be wound on a circular cylinder of radius

$$\begin{aligned} r' &= \rho \sin^2 \psi \\ &= \frac{mv \sin \psi}{eH}. \end{aligned}$$

Now let the field, instead of being uniform be that due to a single pole of strength M, so that at a distance r from the pole the strength of the field is M/r^2. It has been shown by Poincaré [2] that in such a case the particle will move along a geodetic line on the cone of revolution of which the generator is the line joining the particle to the pole and semi-apical angle is given by $\sin \omega = r'/r = mvr \sin \psi/Me$. The quantity $r \sin \psi$ is thus constant being equal to $\sin \omega \, eM/mv$. If, therefore, we draw a line from the apex parallel to the line indicating the direction of motion of the particle when it is at a distance r from the pole, then d, the distance between the two lines, will remain constant throughout the motion of the particle, since $d = r \sin \psi$. As the particle moves towards the apex, r decreases and hence ψ the angle between the trajectory and the line of force will increase. When $r = d$, ψ becomes equal to 90° and the turning point of the trajectory is reached. Here the trajectory is nearly of circular form. Further, since the distance

d of the turning point from the apex depends only on the angle ψ at the distance r and not on the momentum of the particle mv, it follows that if, at a given distance from the pole, particles of different mass and velocity start in the same direction, they will all reach the same minimum distance d from the pole before turning back.

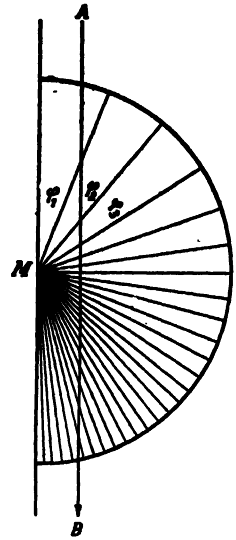


FIG. 2. Illustrating the case when the cone angle is imagined to decrease gradually as the plane rolls; the angles ϕ_1, ϕ_2, ϕ_3 are of gradually diminishing values.

Again ,substituting for $\sin \omega = \phi/2\pi$ and for $\sin \psi = d/r$ in the expression $\sin \omega = mvr \sin \psi/eM$ we find $\pi/\phi = eM/2mvd$, which gives the number of times the orbit turns round the cone on its way from infinity and back to infinity. This number is inversely proportional to the velocity v of the particle.

A very important corollary which considerably simplifies the consideration of the trajectories of the particles when they enter the atmosphere and lose velocity due to encounter with air particles follows from the above. Such a particle, provided its direction of motion is not changed by encounter with air molecules, will, though moving with gradually diminishing velocity, reach the same distance d from the apex as it would have reached if it were moving with constant velocity in empty space. The effect of decrease of velocity is to diminish the value of the cone angle ω and consequently to increase the number of turns. The particle in fact will not be moving on a simple cone of revolution but on a cone whose apical angle may be imagined to be decreasing with each turn. Fig. 2 shows the plane which is developed from a cone of this sort. The angle ϕ diminishes with each revolution of the cone.

8. SOME SOLAR DATA

The sun is a star of the 5th magnitude belonging to the category of the so-called *dwarf stars*. In the spectral sequence of stars it belongs to *Class G*.

The diameter of the sun is $1 \cdot 392 \times 10^6$ km.; the earth-sun distance is $149 \cdot 5 \times 10^6$ km.

Near the centre of the sun the specific gravity is 30, the temperature is several million degrees and the pressure is many hundreds of millions of kilograms per square centimetre.

Period of Rotation.—The sun rotates at different speeds in different latitudes. The mean value of *sidereal rotation* may be taken to be that obtaining at the solar equator, namely 25·35 days.

The average time taken by the sunspots to return to the same position of the disc with respect to the terrestrial observer is, however, 27 terrestrial days. This interval is called the period of *synodic rotation*. It is not the

true period, because the earth moves forward a certain distance in its orbit during the interval of a solar rotation.

The sun's equatorial plane forms an angle of about 7° with the plane of earth's orbit (ecliptic).

The *mean density* of the sun is 1·4 times that of water. Yet the interior of the sun is believed to be in gaseous state. Owing to the enormous pressure in the interior, the atoms are 'crushed' so that the nucleus is shorn of most of the surrounding electrons. Such atoms can come closer together and occupy smaller space. The interior also contains radiation, captivated as it were by the enormous pressure, which slowly forces its way towards the upper surface where it is liberated and goes out into space as heat and light.

Solar atmosphere.—The sun possesses an atmosphere which may be regarded as divided into various layers.

Photosphere.—The bright surface of the sun visible to the naked eye is called photosphere.

Faculae.—The surface of the photosphere examined with telescope shows that it is not uniformly bright as it appears to the naked eye. Examined by spectroheliograph it shows up minute clouds or grains (several hundred kilometres across) alternately bright and dark. These grains or clouds are regions in which currents from the interior of the sun carry hot vapours (hydrogen or calcium) to such heights that condensation occurs. The small bright clouds often run into one another and form bright patches on the surface of the sun called *faculae*. These are often found round the sunspots and like the latter have an 11-year cycle.

Reversing Layer.—Above the photosphere there is a layer consisting of various gases comparatively cooler than the photosphere. This layer, about 600 km. thick, absorbs some of the radiant light emanating from the photosphere and is called the *reversing layer*. This is because it causes dark Fraunhofer lines to appear on the solar spectrum as observed from the earth.

Chromosphere.—Above 600 km. the reversing layer gradually merges into the *chromosphere* which may be regarded as the upper atmosphere of the sun. It extends to several thousands of kilometres and contains the gases which are abundant in the reversing layer, like hydrogen, helium, calcium, etc.

Prominences—Flocculi.—When chromospheric gases—ionized calcium, helium and hydrogen vapours—rise to great heights they are seen in all their grandeur as *prominences* near the limb of the sun. The heights attained by the prominences are enormous—averaging 50,000 km., in extreme cases up to 500,000 km. After attaining the great heights the prominences curve round to fall back on the surface of the sun. The principal emission lines always found in the spectrum of the prominences are those of the Balmer series (hydrogen), ionized calcium (H and K) and helium (D_3).

Projections of the prominences on the solar disc, as recorded by the spectroheliograph, are called bright (or, dark) *flocculi* or *filaments*. (They

are bright if they emit and dark if they absorb their characteristic wave lengths.)

Bright eruptions (solar flare).—Intense bright patches known as *bright eruptions* are sometimes seen on the disc of the sun with the spectroheliograph. The typical bright eruption develops from hydrogen flocculus and occurs in the neighbourhood of sunspots. Spectroheliographic observations do not show the existence of any marked outward velocity of the bright matter of an eruption. The phenomenon may therefore be more properly described as *solar flare*.

Corona.—When the sun's disc is hidden by moon, as during a total solar eclipse (or, by an occulting disc, as in a coronagraph), diffuse illumination of various fantastic shapes—extending to several million kilometres—is seen round the disc. This diffuse light is called *solar corona.*

Two parts of the corona may be distinguished. There is a pale yellow glow extending about a quarter of a solar radius beyond the rim. This is called *inner corona.* White and fainter streamers are also seen to reach out several solar radii beyond the pale yellow glow. This is called *outer corona.*

The coronal spectrum consists of a continuous part superposed by lines. The most intense of the line radiations is the so-called green coronal line (λ 5303) and is due to Fe^{+13}. The continuous spectrum consists of two components—called the K and F components. The K-component is due to light scattered by the atmosphere of free electrons which is supposed to surround the sun. This component is strongly polarized and has no Fraunhofer lines. (They are blurred out by Doppler effect.) The F-component is due to light scattered by inter-planetary dust cloud (the same as produces zodiacal light). This component is unpolarized (or, only, slightly polarized) and shows unbroadened Fraunhofer lines. Near the limb of the sun the intensity of the F-component is much less than that of the K-component. But, the intensity decreases more slowly with increasing distance and exceeds the intensity of the K-component at distances larger than one solar radius.

Sunspots.—On the photosphere are found dark spots visible even to the naked eye. These are centres of vortex motion of gases as is evidenced by spectroheliograms taken with H_α lines. Their number varies from year to year, and they become most numerous at intervals of 11 years (see Fig. 28, Chap. VII, Sec. 7).

The spots are confined to two zones of solar latitude, 5° and 40° north or south of the solar equator. With the progression of the solar cycle the spots, while they decrease in number, move towards the solar equator and almost die out when they have arrived at about latitude 8°. Fresh sunspots at the higher latitude appear two or three years before the final disappearance of the last member of the solar cycle at lower latitudes.

A numerical measure of the sunspot activity is provided by the so-called *Wolf number*. The Wolf sunspot number is obtained, for each day, by multiplying the number of distinct visible sunspot groups by 10 and adding to it the number of individual spots observable in the groups.

Monthly and annual sunspot numbers are obtained from the numbers assigned to the individual days.

The *Zürich sunspot number* is measured in the same way and is called so because it is dependent on observations at Zürich Observatory and its stations at Locarno and Arosa.

Magnetic field.—Study of the Zeeman effect of spectral lines of the light from the sunspots shows the existence of intense magnetic field (about 3,000 gauss). The magnetic field is produced by the motion of the ionized particles in the vortices.

Besides the field on the sunspots the sun is also supposed to have a general magnetic field (about 50 gauss at the surface) with the magnetic axis inclined approximately at an angle of 6° to its axis of rotation. The existence of the general field is supported by the appearance of the corona, specially of the coronal rays near the poles, which is similar to the distribution of the lines of force surrounding a magnetized sphere. However, recent measurements (slight amounts of circular polarization in the wings of spectral lines) seem to cast doubt on the existence of such a field—at least a permanent field [1, 2]. It is possible that if a general magnetic field exists, it is variable and is correlated with the phase of the sunspot cycle.

Solar spectrum—Temperature.—The distribution of energy in the visible portion of the solar spectrum, apart from the dark Fraunhofer lines, roughly corresponds to that of a black body at 6000°K—the maximum energy lying near 5000 A. The spectrum ends rather abruptly at 2900 A (see Fig. 1, p. 119) due to absorption by atmospheric ozone (λ 2900–2200). The absorption due to molecular oxygen commences at about 2100 A. (For further details of these absorptions, see Sec. 2.) There is thus a 'window' in the region λ 2200–2100. Attempts have been made to detect solar radiation at wavelengths around 2100 A, but with doubtful success.

The energy distribution of solar radiation beyond λ 2900 may not correspond at all to that of a black body at 6000°K. Indeed, there are several reasons for believing that the sun radiates in the ultraviolet far more strongly than such a body [see Chapter VI, Sec. 13 (*d*)]. Between λ 1000 and λ 200 the radiation leaving the sun may correspond to that of temperature 20,000°K. From the chromosphere and the prominences there are also line emissions at all times in the principal series of H (L_α —λ 1215, 1026, . . . 912), He (λ 584, 537, . . . 506) and He^+ (λ 304, 256 . . . 228). The intensities of these emissions are increased many thousand times during transitory eruptions (lasting from 20–60 minutes) from limited areas of the sun's disc. The emissions λ 1215 and λ 584 due to H and He respectively as also lines due to Ca^+ are specially intensified.

Energy received by the earth—Solar constant.—The quantity of solar energy received by 1 sq. cm. surface held normally to the solar rays outside the atmosphere at the mean earth-sun distance is called the *solar constant*. Its value is estimated to be about 2·0 cals. cm.$^{-2}$ min.$^{-1}$

Assuming the sun to be radiating like a black body at 6000°K the energy contained in the spectral region beyond any wavelength can be calculated by Planck's radiation formula. Thus, the number of quanta whose energy exceeds a certain limit $h\nu$ is proportional to I_1 where,

$$I_1 = \int_{x_\nu}^{\infty} \frac{x^2}{e^x - 1}\,dx \text{ and } x = \frac{h\nu}{kT}.$$

The number of quanta N_ν within the range $\nu \rightarrow \nu_\infty$ entering the terrestrial atmosphere per second per sq. cm. is then given by

$$N_\nu = 3 \times 10^{17}\, I_1 \text{ (approximately).}$$

N_ν is thus the maximum possible number of quanta available for a photo-reaction whose threshold frequency is ν. In the Table below N_ν;

TABLE

Numbers of quanta available (on the assumption of a hot sun at 6000°K) for important photo-reactions in the upper atmosphere.

Photo-reaction process.	Wavelength corresponding to the threshold frequency ν at which the reaction begins.	Number of quanta N_ν, whose energy exceeds $h\nu$, entering earth's atmosphere per sq. cm. per sec.
Ozone absorption in terrestrial atmosphere	2900 A (4·3 eV)	$6{\cdot}3 \times 10^{15}$
Runge-Schumann absorption .. .. $O_2 + h\nu \rightarrow O_2$ (excited)	1925 A (6·4 eV)	$2{\cdot}1 \times 10^{14}$
Runge-Schumann continuum .. .. $O_2 + h\nu \rightarrow O$ (excited) $+ O$	1760 A (7·0 eV)	$5{\cdot}2 \times 10^{13}$
First ionization of O_2 .. .. .. $O_2 + h\nu \rightarrow O_2^+$ (normal) $+ e$	1019 A (12·2 eV)	$1{\cdot}0 \times 10^{10}$
First ionization of O .. .. .. $O + h\nu \rightarrow O^+$ (normal) $+ e$	910 A (13·5 eV)	$9{\cdot}3 \times 10^{8}$
First ionization of N_2 .. .. .. $N_2 + h\nu \rightarrow N_2^+$ (normal) $+ e$	795 A (15·5 eV)	$2{\cdot}7 \times 10^{7}$
Second ionization of O_2 .. .. $O_2 + h\nu \rightarrow O_2^+$ (excited) $+ e$	770 A (16·1 eV)	$8{\cdot}1 \times 10^{6}$
Third ionization of O_2 .. .. .. $O_2 + h\nu \rightarrow O_2^+$ (excited) $+ e$	732 A (16·9 eV)	$2{\cdot}1 \times 10^{6}$
Second ionization of O .. .. $O + h\nu \rightarrow O^+$ (excited) $+ e$	732 A (16·9 eV)	$2{\cdot}1 \times 10^{6}$
Fourth ionization of O_2 .. .. $O_2 + h\nu \rightarrow O_2^+$ (excited) $+ e$	682 A (18·2 eV)	$1{\cdot}8 \times 10^{5}$
Third ionization of O .. .. .. $O + h\nu \rightarrow O^+$ (excited) $+ e$	665 A (18·5 eV)	$1{\cdot}0 \times 10^{5}$
Second ionization of N_2 .. .. $N_2 + h\nu \rightarrow N_2^+$ (excited) $+ e$	661 A (18·7 eV)	$7{\cdot}4 \times 10^{4}$

is given for a number of important photo-reactions in the upper atmosphere. It should, however, be remembered that the actual numbers of quanta received in the extreme ultraviolet, in all probability, far exceed those given in the Table calculated on the assumption that the sun is radiating like a hot black-body at 6000°K.

9. UPPER ATMOSPHERIC NOMENCLATURE

In this Section we shall list a number of terms defining the different upper atmospheric regions and their characteristics. The nomenclature used in the list is based principally on the comprehensive recommendations made by a Panel appointed by the Special Sub-Committee on the Upper Atmosphere of the National Advisory Committee for Aeronautics of the U.S.A. (NACA). Two kinds of nomenclature have been proposed: (*a*) A general nomenclature covering the whole range of heights in the atmosphere and (*b*) different special nomenclatures, each based on a particular characteristic of the atmosphere (e.g. temperature, ionization as is of importance to radio propagation, etc.). [For explanation of the asterisk (*) and double asterisk (**) marks against some of the terms see the note below.]

The recommended nomenclature could not be utilized in the present volume as it was received too late.

It may be put on record that the first suggestions for the introduction of upper atmospheric nomenclature came from H. Flohn and R. Penndorf [1] and from S. Chapman [2]. The panel appointed by the Special Sub-Committee on Upper Atmosphere of the NACA to consider suitable upper atmospheric nomenclature took into consideration the suggestions made by these authors, as also those made by J. Kaplan, N. C. Gerson and A. H. Waynick. The nomenclature, before final adoption, was circulated for eliciting opinions of workers on the upper atmosphere.

The recommendations of the Panel were placed before a Committee on Nomenclature (consisting of members from the Associations of Meteorology and of Terrestrial Magnetism and Electricity) at the August, 1951, meeting of the International Union of Geodesy and Geophysics (U.I.G.G.) held in Brussels. The terms which the Committee viewed with reserve are marked with single asterisk (*), and, those regarding which no recommendation was made concerning their use are marked with double asterisk (**). The rest are recommended for trial use. It has, however, been thought fit to retain the complete list as originally recommended by the NACA Panel on account of its comprehensive nature.

(*a*) General Nomenclature

(1) *Troposphere*, 0–12 km.

(2) *Stratosphere*—the region between (1) and (4), 12–80 km. (This is according to the recommendation of the Panel; the defining characteristics are explained below under Special Nomenclature. The Committee (*vide* note above), however, recommends the retention of the usual sense of the term.)

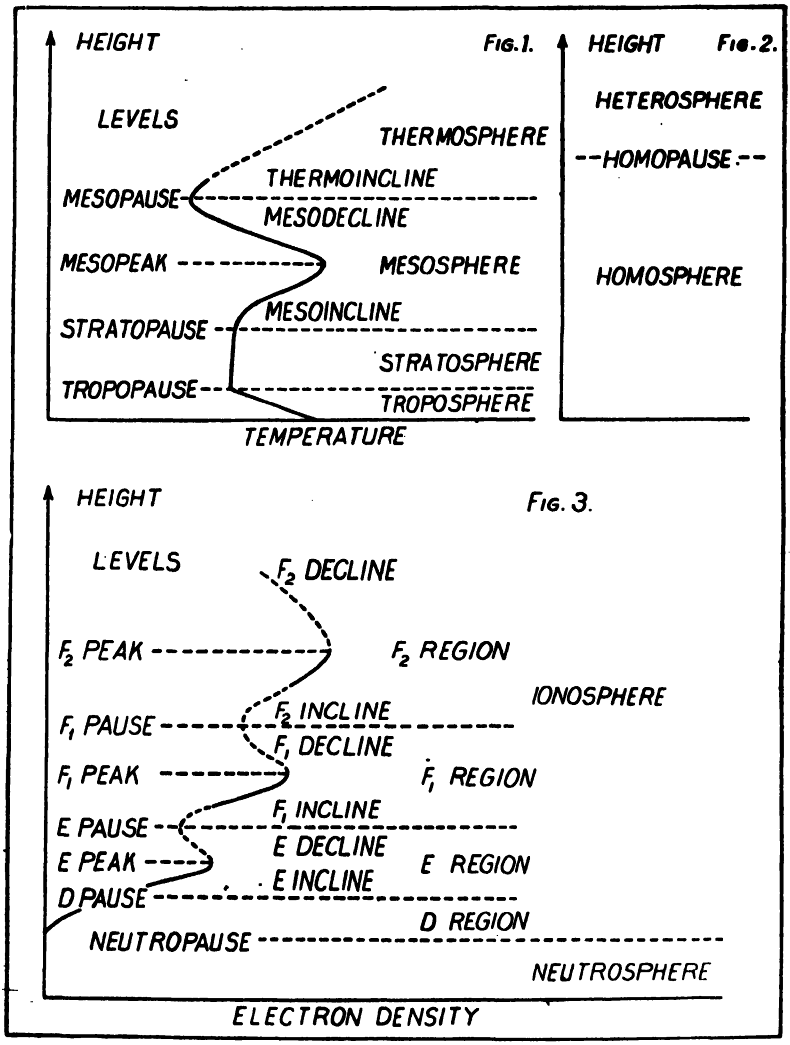
(3) *Mesosphere*—the region between the top of the stratosphere (usual sense) and 80 km., the region of the major minimum

of temperature below 100 km. (This name is not included in the list of the Panel. It was first suggested by Chapman and is recommended by the Committee for trial use.)

(4) *Ionosphere*, 80–400 km.

(5) *Suprasphere* (*)—the region between (3) and (6), 400–1000 km.

(6) *Exosphere*—the region in which there is appreciable escape of air to outer space, above 1000 km.



FIGS. 1, 2, 3. Illustrating upper atmospheric nomenclature as originally proposed by Chapman. See text.

(*b*) Special Nomenclature

(i) Special nomenclature based on temperature distribution:

(1) *Troposphere*—As used in the usual sense.

(2) *Stratosphere*—The region extending from the tropopause to the level of minimum temperature (somewhat below 100 km.), with the use, where convenient, of the name *isosphere* (*) for the lower, nearly isothermal, part of the stratosphere where such layer exists. (See, however, the note above.)

(3) *Thermosphere*—The region of increasing temperature above the stratosphere as defined above.

(ii) Special nomenclature based on composition:

(1) *Turbosphere* (*)—The region from the ground upwards, in which turbulent mixing overcomes the tendency for the various constituents of air to diffuse relative to one another towards a distribution according to their molecular weights.

(2) *Homosphere*—The region in which the mean molecular weight of the air is approximately uniform despite the existence of layers of important small amounts of constituents like ozone so that the scale height in this region depends substantially on the values of gravity and temperature.

(3) *Heterosphere*—The region in which the mean molecular weight differs appreciably from that in the troposphere, whether as a result of dissociation of oxygen or other constituents or, through diffusive separation.

(iii) Special nomenclature based on ionization:

(1) *Ionosphere*—The region in which the atmosphere contains electrons and ions in numbers sufficient to influence radio wave propagation.

(2) *Neutrosphere* (*)—The region in which ions and electrons, though present, are not of importance to radio wave propagation on account of their insufficient numbers. (The name is complementary to ionosphere and obviously signifies the atmosphere below the *D*-region.)

(*c*) Upper boundaries of the layers

The upper boundaries of the various layers or regions are designated by extended use of the term *pause*, as already customary in tropopause.

Layer	Upper boundary
Troposphere	*Tropopause*
Stratosphere	*Stratopause*
Isosphere (*)	*Isopause* (*)
Mesosphere	*Mesopause*
Neutrosphere (*)	*Neutropause* (*)
Homosphere	*Homopause*
Turbosphere (*)	*Turbopause* (*)

(*d*) Levels of maxima of the defining characteristics of the different layers (**)

The levels of maxima of the defining characteristics of the layers are to be signified by the use of the term *peak.*

Stratopeak	..	The level of maximum temperature between tropopause and stratopause.
Ozonepeak—relative (or *absolute*)	..	The level at which the relative (or, absolute) ozone concentration attains its maximum.
E-, *F*-*peak*	..	The level at which the electron number density in the *E*- (or *F*-) region attains its maximum value.

(*e*) Incline and Decline (**)

The regions in which the defining characteristic of a layer is increasing or decreasing in value (with height) are to be signified by the use of the terms *incline* and *decline* respectively. Thus:

Stratoincline	*Stratodecline*
Ozone incline	*Ozone decline*
E (or *F*) *incline*	*E* (or *F*) *decline*

(*f*) Chemosphere

In discussions on upper atmospheric chemistry the term *chemosphere* is to be used for the region of the atmosphere in which chemical reactions are important owing to dissociative processes induced by solar radiation—wave or corpuscle.

BIBLIOGRAPHY

CHAPTER I

General Considerations

1. Milne, E. A., *Trans. Camb. Phil. Soc.*, **22**, 483 (1922-23).
2. Maris, H. B., *Terr. Mag. and Atmos. Elec.*, **33**, 293 (1928); **34**, 45 (1929).
3. Epstein, P. S., *Ger. Beitr. z. Geophys.*, **35**, 153 (1932).
4. Mitra, S. K. and Rakshit, H., *Ind. Jour. Phys.*, **12**, 47 (1938).
5. Jones, J. E., *Trans. Camb. Phil. Soc.*, **22**, 535 (1922-23).
6. Jeans, J. H., *The Dynamical Theory of Gases*, Cambridge University Press, Second Edition, p. 357.

6a. *Ibid.*, p. 252.

6b. Chackett, K. F., Paneth, F. A., Wilson, E. J., *Jour. Atmos. and Terr. Phys.*, **1**, 49 (1950).

7. Lindemann, F. A., *Quar. Jour. Roy. Met. Soc.*, **65**, 330 (1939).

7a. Spitzer, L., *The Atmospheres of the Earth and Planets*, edited by G. P. Kuiper Chicago University Press, pp. 243–249 (1948).

8. Helge-Petersen, *Pub. Danske Meteorologiske Institut (Copenhagen)*, No. 6 (1928).
9. Mitra, S. K. and Banerjee, A. K., *Ind. Jour. Phys.*, **13**, 107 (1939).
10. Jones, Sir H. S., *Endeavour*, **4**, 35 (1945).
11. Chapman, S. and Milne, E. A., *Quar. Jour. Roy. Met. Soc.*, **46**, 392 (1920); Eddington, A. S., *The Internal Constitution of the Stars*, Cambridge University Press (1926), p. 281.
12. Stőrmer, C., *University Observatory, Oslo*, Publication No. 10, 1934.
13. Stőrmer, C., *Quatrième Communication, Vid. Selsk. Skrifter, Math.-Naturv. Klasse*, No. 12 (1916).

CHAPTER II

Atmospheric Oscillations—Tides in the High Atmosphere

1. Laplace, P. S., *Mécanique Céleste*, Chap. 5. An account of the literature up to 1924 is given by Chapman, S., *Quar. Jour. Roy. Met. Soc.*, **50**, 166 (1924).
2. Bennett, W. J., *Monthly Weather Review*, **34**, 528 (1906).
3. Simpson, G. C., *Quar. Jour. Roy. Met. Soc.*, **44**, 1 (1928).
4. Wilkes, M. V., *Oscillations of the Earth's Atmosphere*, Cambridge University Press (1949), p. 14.

4a. Chapman, S., Address to International Association of Meteorology, International Union of Geodesy and Geophysics, Oslo, 1948.

5. Adolf Schmidt, *Handbuch der Experimentalphysik, Geophysik*, Bd. **25**, Teil 1, p. 176.

5a. Chapman, S., *Quar. Jour. Roy. Met. Soc.*, **50**, 166 (1924).

6. Kelvin, Lord (Thomson, Sir William), *Proc. Roy. Soc. Edin.*, **7**, 396 (1882).
7. Lamb, H., *Proc. Roy. Soc. A.*, **84**, 551 (1911).
8. Taylor, G. I., *Proc. Roy. Soc. A*, **126**, 169 (1929).
9. Strachey, Lieut. Gen., *The Eruption of the Krakatau*—Report of the Krakatau Committee of the Royal Society (1888).
10. Whipple, F. J. W., *Quar. Jour. Roy. Met. Soc.*, **56**, 287 (1930).
11. Taylor, G. I., *Proc. Roy. Soc. A*, **156**, 318 (1936).
12. Pekeris, C. L., *Proc. Roy. Soc. A*, **158**, 650 (1937).

13. Weekes, K. and Wilkes, M. V., *Proc. Roy. Soc. A*, **192**, 80 (1947).
14. Wilkes, M. V., *Oscillations of the Earth's Atmosphere*, Cambridge University Press (1949), p. 60-61.
15. Sabine, E., *Phil. Trans. Roy. Soc. A.*, **137**, 45 (1847).
16. Chapman, S., *Quar. Jour. Roy. Met. Soc.*, **44**, 271 (1918); *Presidential Address, Jour. Lond. Math. Soc.*, **7**, 68 (1931); *Proc. Roy. Soc. A*, **151**, 105 (1935).
17. Chapman, S., *Presidential Address International Union of Geodesy and Geophysics* (1939).
18. Chapman, S. and Tschu, K. K., *Proc. Roy. Soc. A*, **195**, 310 (1948).
19. Chapman, S., Pramanik, S. K., and Topping, J., *Ger. Beitr. z. Geophys.*, **33**, 246 (1931).
20. Chapman, S., *Observatory*, **60**, 154 (1937).

CHAPTER IIIA

Anomalous Propagation of Sound

1. Borne, G. von dem, *Die Erdbebenwarte*, **4**, 1 (1904); **6**, 110 (1906).
2. Whipple, F. J. W., *Nature*, **111**, 187 (1923).
3. Gowan, E. H., *Proc. Roy. Soc. A*, **128**, 531 (1930); *Quar. Jour. Roy. Met. Soc., Suppt.*, **62**, 34 (1936).
4. Schrödinger, E., *Phys. Zeits.*, **18**, 445 (1917).
5. Emden, R., *Met. Zeit.*, **35**, 13, 74 and 114 (1918).
6. Meisser, O., *Handbuch der Experimentalphysik*, Bd. **25** III, Kap. I, 232 (1930).
7. Tucker, W. S. and Paris, E. T., *Phil. Trans. Roy. Soc. A*, **221**, 389 (1921).
8. Benioff, H. and Gutenberg, B., *Nature*, **144**, 478 (1939).
9. Meisser, O., *Zeits. Geophys.*, **3**, 287 (1927).
10. Whipple, F. J. W., *Quar. Jour. Roy. Met. Soc.*, **61**, 285 (1935).
10a. Mathur, L. S., *Ind. Jour. Met. and Geoph.*, **1**, 24 (1950).
11. Gutenberg, B., *Zeits. Geophys.*, **3**, 260 (1926); *Ger. Beitr. z. Geophys.*, **27**, 217 (1930).
12. Gutenberg, B., *Bull. Am. Met. Soc.*, **20**, 192 (1939).
13. Batesman, H., *Phil. Mag.*, **19**, 576 (1910); Herlotz, *Phys. Zeits.*, **8**, 145 (1907).

CHAPTER IIIB

Meteoric Phenomena

1. Porter, J. G., *Rep. Prog. Phys.*, Physical Society, London, **11**, 409 (1946-47).
2. Whipple, F. L., *Proc. Amer. Phil. Soc.*, **83**, 711 (1940).
3. Hoffmeister, C., *Die Meteore, Leipzig* (1937); *Meteorströme*, Weimar (1948).
4. Porter, J. G., *Mem. Brit. Astr. Assc.*, **34**, 37 (1943); *Mon. Not. Roy. Astr. Soc.*, **103**, 134 (1943); *ibid.*, **194**, 257 (1943).
5. McIntosh, R. A., *Trans. Roy. Soc. N.Z.*, **66**, 60 (1936); *ibid.*, **69**, 392 (1940).
6. Olivier, C. P., *Meteors, Published by Williams and Welkins, Baltimore* (1938); Prentice, J. P. M., *British Astr. Assocn. Handbook*, p. 36 (1945); *Rep. Prog. Phys.*, Physical Society, London, **11**, 389 (1946-47).
7. Öpik, E., *Publ. Observ. Astron. Univ. Tartu*, **30**, No. 6 (1941).
8. Whipple, F. L., *Proc. Amer. Phil. Soc.*, **79**, 499 (1938); **82**, 275 (1940); **83**, 711 (1940); *Sky and Telescope*, **8**, 1 (1949).
9. Whipple, F. L., *Rev. Mod. Phys.*, **15**, 246 (1943).
10. Whipple, F. L., *Sky and Telescope*, **8**, 1 (1949).
11. Öpik, E., *Mon. Not. Roy. Astr. Soc.*, **100**, 315 (1940).
12. McIntosh, R. A., *Mon. Not. Roy. Astr. Soc.*, **100**, 510 (1940).

13. Sparrow, C. M., *Astrophys. Jour.*, **63**, 90 (1926).
14. Epstein, P., *Proc. Nat. Acad. Sci.*, **17**, 539 (1931).
15. Lindemann, F. A., and Dobson, G. M. B., *Proc. Roy. Soc. A*, **102**, 411 (1923); Lindemann, F. A., *Astrophys. Jour.*, **65**, 117 (1929).
16. Herlofson, N., *Rep. Prog. Phys.*, Physical Society, London, **11**, 444 (1946-47).
17. Öpik, E., *Acta Univ. Tartu, A*, 26: *Harvard Reprints No. 100* (1933); *Publ. Observ. Astron. Univ. Tartu*, **24**, No. 5 (1937); *Mon. Not. Roy. Astr. Soc.*, **100**, 315 (1940).
18. Compton, K. T. and Langmuir, J., *Rev. Mod. Phys.*, **2**, 186 (1930).
19. Van Voorhis, C. C. and Compton, K. T., *Phys. Rev.*, **37**, 1596 (1931).
20. Hoffleit, D., *Proc. Nat. Acad. Sci.*, **19**, 212 (1938).
21. Hoffmeister, C. F., *Die Meteore*, Leipzig (1937).
22. Öpik, E., *Harvard Reprints No. 100* (1933).
23. Öpik, E., *Publ. Observ. Astron. Univ. Tartu*, **24**, No. 5 (1937).
23a. Jacchia, G., *Harvard College Observatory and Centre of Analysis of M.I.T. Technical Report No. 3* (1949).
24. Hoppe, J., *Astronom. Nach.*, **262**, 160 (1937).
24a. Jacchia, G., *Harvard College Observatory and Centre of Analysis of M.I.T., Technical Report No. 4* (1949).
25. Skellett, A. M., *Phys. Rev.*, **37**, 1668 (1931).
26. Schafer, J. P. and Goodall, W. M., *Proc. I.R.E.*, **19**, 1434 (1931); **20**, 1131 (1932); **20**, 1941 (1932).
27. Mitra, S. K., Syam, P. and Ghose, B. N., *Nature*, **133**, 533 (1934).
28. Bhar, J. N., *Ind. Jour. Phys.*, **12**, 109 (1937).
29. Minohara, T. and Ito, Y., *Rep. Rad. Res. (Japan)*, **3**, 115 (1933).
30. Pickard, G. W., *Proc. I.R.E.*, **19**, 1166 (1931).
31. Lovell, A. C. B., *Rep. Prog. Phys.*, Physical Society, London, **11**, 415 (1948).
32. Manning, L. A., *Jour. Appl. Phys.*, **19**, 689 (1948).
33. McKinley, D. W. R. and Millman, P. M., *Proc. I.R.E.*, **37**, 364 (1949).
33a. Huruhata, A., *Pub. Astr. Soc., Japan*, **1**, 39 (1949).
34. Pierce, J. A., *Proc. I.R.E.*, **26**, 892 (1938); *Phys. Rev.*, **59**, 625 (1941); *ibid.*, **71**, 88 (1947).
35. Blackett, P. M. S. and Lovell, A. C. B., *Proc. Roy. Soc. A*, **177**, 183 (1941).
36. Lovell, A. C. B. and Clegg, J. A., *Proc. Phys., Soc., London*, **60**, 491 (1948); *Nature*, **160**, 670 (1947).
37. Hey, J. S. and Stewart, G. S., *Proc. Phys. Soc. A, London*, **59**, 858 (1947).
38. Herlofson, N., *Observatory*, **68**, 220 (1948).
39. Lovell, A. C. B., *Sci. Progress*, **38**, 22 (1950).
39a. McKinley, D. W. R. and Millman, P. M., *Proc. I.R.E.*, **37**, 364 (1949).
39b. Hey, J. S., Parsons, S. J. and Stewart, G. S., *Mon. Not. Roy. Astr. Soc.*, **107**, 176 (1947).
40. Ratcliffe, J. A., *Nature*, **162**, 9 (1948).
41. Herlofson, N., *Arkiv f. Fysik*, Band 3, nr 15, p. 247 (1951).
42. Tonks, L. and Langmuir, I., *Phys. Rev.*, **33**, 195 and 990 (1929).
42a. Romell, D., *Nature*, **167**, 243 (1951).
43. Clegg, J. A. and Davidson, I. A., *Phil. Mag., Ser. 7*, 41, 77 (1950).
44. Taylor, D. and Westcott, C. H., *Jour. I.E.E., Pt. IIIA*, **93**, 588 (1946).
45. Clegg, J. A., *Phil. Mag., Ser. 7*, 39, 577 (1948).
46. Clegg, J. A., Hughes, V. A., Lovell, A. C. B., *Mon. Not. Roy. Astr. Soc.*, **107**, 369 (1947).
47. Clegg, J. A., *Jour. Brit. Astr. Soc.*, **58**, No. 7, 271 (1948).
48. Aspinall, A., Clegg, J. A., Lovell, A. C. B., *Mon. Not. Roy. Astr. Soc.*, **109**, 352 (1949).
49. Hey, J. S., *Nature*, **159**, 119 (1947).

50. Hey, J. S., Parsons, S. J., and Stewart, G. S., *Mon. Not. Roy. Astr. Soc.*, **107**, 176 (1947).
51. Millman, P. M. and McKinley, D. W. R., *J. Roy. Astronom. Soc., Canada*, **42**, 121 (1947).
52. Millman, P. M. and McKinley, D. W. R., *Canadian Jour. Research*, **27**, 53 (1949).
53. Davies, J. G. and Ellyett, C. D., *Phil. Mag., Ser. 7*, 40 (1949).
54. Chamanlal and Venkataraman, K., *Electro-techniques (Bangalore)*, **14**, 23 (1941).
55. Manning, L. A., Villard, O. G., and Peterson, A. M., *Jour. Appl. Phys.*, **20**, 475 (1949).
56. Appleton, E. V. and Naismith, R., *Proc. Phys. Soc. (London)*, **59**, 461 (1947).
57. McKinley, D. W. R., Private communication to Lovell, A. C. B.; *Sci. Prog.*, **38**, 36 (1950).

CHAPTER IV

OZONOSPHERE

1. Abbot, C. G., *Smithsonian Solar Radiation Researches, Ger. Beitr.*, **16**, 344 (1927); *Ann. Obs. Smithson. Inst.*, **3**, 200 (1913).
2. Hartley, W. N., *Jour. Chem. Soc.*, **39**, 57 and 111 (1881).
3. Huggins, M. L., *Proc. Roy. Soc. A*, **48**, 216 (1890).
4. Fowler, A. and Strutt, Hon. R. J., *Proc. Roy. Soc. A*, **93**, 577 (1917).
5. Strutt, Hon. R. J., *Proc. Roy. Soc. A*, **94**, 260 (1918).
6. Götz, F. W. P., Meetham, A. R., and Dobson, G. M. B., *Proc. Roy. Soc. A*, **145**, 416 (1934).
7. Regener, E. and Regener, V. H., *Phys. Zeit.*, 35, 788 (1934).
8. *Rep. Prog. Phys.*, Physical Society, London, 9 (1942-43); Articles by Price, W. C. p. 10; Bamford, C. H., p. 75; Chapman, S., p. 92.
9a. Fabry, C. and Buisson, H., *Jour. de Phys.*, **3**, 196 (1913).
9b. Läuchli, A., *Zeit. Phys.*, **53**, 92 (1929).
10. Ny Tsi-Zé and Choong Shin-Piaw, *C. R. Acad. Sci. (Paris)*, **195**, 309 (1932); **196**, 916 (1933); *Chinese Jour. Phys.*, **1**, 1 (1933).
11. Chalonge, D. and Lefèbre, L., *C. R. Acad. Sci. (Paris)*, **197**, 444 (1933).
12. Colange, G., *Jour. de Phys.*, **8**, 254 (1927).
13. Maris, H. B., *Terr. Mag. and Atmos. Elec.*, **33**, 233 (1928); **34**, 45 (1929).
14. Hettner, G., Pohlmann, R., and Schumacher, H. J., *Zeit. Phys.*, **91**, 372 (1934).
15. Wulf, O. R. and Malvin, E. H., *Phys. Rev.*, **38**, 330 (1931).
16. Vassy, E., *Bull. Fran. de Phys.*, No. 402, p. 50 (1937). (See also Conference on Atmospheric Ozone (Oxford), *Quar. Jour. Roy. Met. Soc. Suppt.*, **62**, 32 (1936).)
17. Regener, V. H., *Nature*, **167**, 276 (1951):
18. Fabry, C. and Buisson, H., *Jour. de Phys.*, **2**, 297 (1921).
19. Dobson, G. M. B., *Proc. Phys. Soc. (London)*, **43**, 324 (1931).
20. Regener, E., *Beitr. Phys. frei. Atmos.*, **22**, 249 (1935). Regener, E. and Regener, V. H., *Nature*, **134**, 380 (1934).
21. Götz, F. W. P., *Strahlenther*, **40**, 690 (1931).
21a. Chiplonkar, M. W., *Current Science (Bangalore)*, **8**, 312 (1939).
22. Dobson, G. M. B. and Harrison, D. N., *Proc. Roy. Soc. A*, **110**, 660 (1926).
23. Götz, F. W. P., *Ger. Beitr.*, **31**, 119 (1931).
23a. Strong, J., *Jour. Franklin Inst.*, **232**, 1 (1941).
24. Warburg, E., *Z. Elektrochem.*, **27**, 133 (1921).
25. Ladenburg, R. and Van Voorhis, C. C., *Phys. Rev.*, **43**, 315 (1933).
26. Flory, P. J., *Jour. Chem. Phys.*, **4**, 23 (1936).
27. Kreusler, H., *Ann. d. Phys.*, **6**, 418 (1901).
28. Saha, M. N., *Proc. Roy. Soc. A*, **160**, 155 (1937); *Proc. Nat. Inst. Sci. (India)*, **1**, 217 (1935).

28a. Warburg, E., *Sitzb. kgl. preuss. Akad*, p. 644 (1913).
28b. Forbes, G. S. and Heidt, L. J., *Jour. Amer. Chem. Soc.*, **56**, 1671 (1934).
29. Chapman, S., *Mem. Roy. Met. Soc.*, **3**, 103 (1930).
30. Kistiakowsky, G. B., *Zeit. Phys. Chem.*, **117**, 337 (1925).
31. Mecke, R., *Trans. Faraday Soc.*, **27**, 375 (1931); *Zeit. Phys. Chem.*, Bodenstein Festband, p. 392 (1931).
32. Wulf, O. R. and Deming, L. S., *Terr. Mag. and Atmos. Elec.*, **42**, 195 (1937); **41**, 299 (1936).
33. Eucken, A. and Patat, F., *Zeit. Phys. Chem. B*, **33**, 459 (1936).
34. Gowan, E. H., *Proc. Roy. Soc. A*, *120*, 655 (1928); *128*, 531 (1930); Conference on Atmospheric Ozone (Oxford) *Quar. Jour. Roy. Met. Soc. Suppt.*, **62**, 34 (1936).
34a. Gowan, E. H., *Proc. Roy. Soc. A*, **190**, 219 (1947).
34b. Gowan, E. H., *Proc. Roy. Soc. A*, **190**, 227 (1947).
34c. Strong, J., *Phys. Rev.*, **55**, 1114 (1939).
34d. Summerfield, Doc. Thesis. Cal. Tech. (1941).
34e. Elsasser, W. M., *Harv. Met. Stud. No. 6.*, p. 96 (1942).
35. Penndorf, R., *Beiträge zum Ozonproblem* (*Leipzig*, 1936), p. 257.
35a. Penndorf, R., *Ann. Geophys.*, **6**, 4 (1950).
36. Fabry, C. and Buisson, H., *Jour. de Phys.*, **3**, 196 (1913).
37. Hettner, G., *Ann. d. Phys.*, **55**, 476 (1918).
38. Gold, E., *Proc. Roy. Soc. A*, **82**, 43 (1909).
39. Jahnke, E. and Emde, F., *Funktionentafeln, Leipzig* (B. C. Teubner), 1909.
40. Dobson, G. M. B., *Proc. Roy. Soc. A*, **129**, 411 (1930).
41. Barbier, D., Chalonge, D., and Vassy, E., *Rev. Opt.*, **13**, 199 (1934).
42. Vassy, E., *C. R. Acad. Sci.* (*Paris*), **202**, 1426 (1936).
43. Gauzit, J., *Ann. d. Phys.*, **4**, 450 (1935).
43a. Karandikar, R. V., *Proc. Ind. Acad. Sci.*, **28**, 63 (1948).
43b. Vassy, E. Mons. and Vassy, E. Mme., *C. R. Acad. Sci.*, Paris, **228**, 764 (1949).
44. Chiplonkar, M. W., *Proc. Ind. Acad. Sci.* (*Bangalore*), **10**, 105 (1939).
45. Lejay, Rev. P., *Notes de Météorologie Physique*, 1937 Fascicule VII.
46. Mitra, S. K., *Science and Culture* (*Calcutta*), **3**, 496 (1937-38).
47. Chapman, S., *Phil. Mag.*, **10**, 345 (1930).
47a. Malurkar, S. L., *Science and Culture* (*Calcutta*), **16**, 70 (1951).
48. Dobson, G. M. B., Harrison, D. N., and Lawrence, J., *Proc. Roy. Soc. A*, **122**, 456 (1929); Also Dobson, G. M. B., Brewer, A. W., and Cwilong, B. M., *Proc. Roy. Soc. A*, **185**, 144 (1946).
49. Tonsberg, E. and Chalonge, D., Conference on Atmospheric Ozone (Oxford), *Quar. Jour. Roy. Met. Soc. Suppt.*, **62**, 55 (1936).
50. Meetham, A. R., Conference on Atmospheric Ozone (Oxford), *Quar. Jour. Roy. Met. Soc. Suppt.*, **62**, 59 (1936).
51. Cabannes, J. and Dufay, J., *Jour. de Phys.*, **8**, 353 (1927).
52. Fowle, F. E., *Smithson. Misc. Coll.*, **81**, No. 11 (1929).
53. Fowle, F. E., *Trans. Am. Geoph. Union*, p. 160 (1934).
54. Götz, F. W. P., *Conseil International des Unions Scientifiques, 6e Rapport Comm.* Relations éntre les Phénomènes Solaires et Terrestres, Orléans (1948).
55. Götz, F. W. P., Périodicités dans les phénomènes de la haute atmosphère, *Conf. Intern.*, *Lyon* (1947).
56. Mitra, S. K., *Zbl. Geophysik, Meteor. Geod.*, **4**, 337 (1939).
57. Götz, F. W. P., *Erg. Kosm. Physik*, **3**, 253 (1938).
58. Penndorf, R., *Zeit. f. Meteor.*, **1**, 345 (1947).
59. Penndorf, R., *Jour. Meteor.*, **5**, 152 (1948).
60. Störmer, C., University Observatory, Oslo, Publication No. 6 (1933).
61. Barbier, D., Chalonge, D. and Vigroux, E., *Ann. Astrophys.*, **5**, 1 (1942).
62. Ghosh, S. P., *M.Sc. Thesis* (*unpublished*), Calcutta University (1938).
63. Adel, A., *Astroph. Jour.*, **105**, 406 (1947).

CHAPTER V

Oxygen and Nitrogen in the Upper Atmosphere

1. Herzberg, G., *Astrophys. Jour.*, **89**, 290 (1939).
2. Nicolet, M., *Institut Royal Météorologique de Belgique, Memoires*, Vol. XIX, Contribution à l' étude de la structure de l'ionosphère (1946).
3. Wulf, O. R., and Deming, L. S., *Terr. Mag. and Atmos. Elec.*, **43**, 283 (1938).
4. Penndorf, R., *Jour. Geoph. Res.*, **54**, 7 (1949).
5. Massey, H. S. W., *Negative Ions*, Cambridge University Press, 1938, p. 74.
6. Saha, M. N., *Phil. Mag.*, **40**, 472 (1920).
7. Woltjer, J., *Physica*, **5**, 406 (1925).
8. Pannekoek, A., *Proc. Amsterd. Acad.*, **29**, 1165 (1926).
9. Majumdar, R. C., *Ind. Jour. Phys.*, **12**, 75 (1938).
10. Rakshit, H., *Ind. Jour. Phys.*, **21**, 57 (1947).
10a. Moses, H. E. and Ta-You Wu, *Phys. Rev.*, **83**, 109 (1951).
11. Ladenburg, R. and Van Voorhis, C. C., *Phys. Rev.*, **43**, 315 (1933).
12. Martyn, D. F. and Pulley, O. O., *Proc. Roy. Soc. A*, **154**, 455 (1936).
13. Mitra, S. K. and Rakshit, H., *Ind. Jour. Phys.*, **12**, 47 (1938).
14. Chapman, S., *Proc. Roy. Soc. A*, **132**, 353 (1931).
15. See, for example, *The Atmospheres of the Earth and Planets* (edited by G. P. Kuiper, University of Chicago Press, 1947) discussions in the articles by P. Swings (p. 193) and by L. Spitzer, Jr. (p. 219).
16. Bates, D. R., *Phys. Rev.*, **78**, 492 (1950).
17. Mitra, S. K., *Nature*, **167**, 897 (1951).
18. Gaydon, A. G., *Nature*, **153**, 407 (1944).
19. Dufay, J., *Ann. d' ap.*, **6**, 81 (1943).
20. Gauzit, J., *Cahiers de Phys.*, No. 9, p. 47 (1942).
21. Bernard, R., *Phys. Rev.*, **55**, 511 (1939); *Ann. d' ap.*, **4**, 13 (1941).
22. Deb, S., (In course of publication).

CHAPTER VI

Ionosphere

1. Macdonald, H. M., *Proc. Roy. Soc. A*, **71**, 251 (1902-1903); **72**, 59 (1903); **98**, 409 (1921).
2. Rayleigh, Lord (J. W. Strutt), *Proc. Roy. Soc. A*, **72**, 40 (1903); *Jahr. d. Drahtl. Tel.*, **3**, 445 (1910).
3. Poincaré, H., *Proc. Roy. Soc. A*, **72**, 42 (1903); *Rendiconti di Palermo*, **29**, 169, (1910); *C. R. Acad. Sci. (Paris)*, **54**, 795 (1912).
4. Love, A. E. H., *Phil. Trans. Roy. Soc. A*, **215**, 105 (1915).
5. Watson, G. N., *Proc. Roy. Soc. A*, **95**, 83 and 546 (1918-19).
6. van der Pol, B., *Phil. Mag.*, **38**, 365 (1919).
7. Kennelly, A. E., *Elec. World and Eng.*, **15**, 473 (1902).
8. Heaviside, O., *Ency. Brit. (Ninth Edition)*, **33**, 215 (1902).
9. Balfour Stewart, *Ency. Brit. (Ninth Edition)*, **16**, 181 (1902).
10. Eccles, W. H., *Proc. Roy. Soc. A*, **87**, 79 (1912).
11. Larmor, J., *Phil. Mag.*, **48**, 1025 (1924).
12. Appleton, E. V. and Barnett, M. A. F., *Nature*, **115**, 333 (1925); *Proc. Roy. Soc. A*, **109**, 621 (1925).
13. Smith-Rose, R. L. and Barfield, R. H., *Proc. Roy. Soc. A*, **110**, 580 (1926); **116**, 682 (1927).
14. Breit, G. and Tuve, M., *Phys. Rev.*, **28**, 554 (1926).
15. Appleton, E. V., *Proc. Phys. Soc. (London)*, **37**, 16D (1925); *Jour. I.E.E.*, **71**, 642 (1932).

16. Appleton, E. V., *U.R.S.I. Reports, Washington* (1927).
17. Goldstein, S., *Proc. Roy. Soc. A*, **121**, 260 (1928).
18. Hartree, D. R., *Proc. Camb. Phil. Soc.*, **25**, 47 (1929).
19. Darwin, C. G., *Proc. Roy. Soc. A*, **146**, 17 (1934); **182**, 152 (1943).
20. Ratcliffe, J. A., *Wir. Eng. and Exp. Wir.*, **10**, 354 (1933).
21. Bhar, J. N., *Science and Culture (Calcutta)*, **2**, 322 (1936-37).
22. Mitra, S. K., *Science and Culture (Calcutta)*, **2**, 322 (1936-37).
22a. Aden, A. L., Bettencourt, J. T. de and Waterman, A. T., *Jour. Geoph. Res.*, **55**, 53 (1950).
23. Taylor, Mary, *Proc. Phys. Soc. (London)*, **45**, 245 (1933); **46**, 408 (1934).
24. Goubau, G., *Hochfrequenztech. u. Electroakust.*, **45**, 179 (1935).
24a. Westfold, K. C., *Jour. Atmos. Terr. Phys.*, **1**, 82 (1950).
25. Bailey, V. A., *Phil. Mag.*, **18**, 516 (1934).
26. Martyn, D. F., *Phil. Mag.*, **19**, 376 (1935).
27. Ghosh, S. P., *Ind. Jour. Phys.*, **12**, 341 (1938).
28. Booker, H. G., *Proc. Roy. Soc. A*, **150**, 267 (1935).
29. Farmer, F. T., and Ratcliffe, J. A., *Proc. Roy. Soc. A*, **151**, 370 (1935).
30. Baker, W. G., and Green, A. L., *Proc. I.R.E.*, **21**, 1103 (1933).
31. Booker, H. G., *Phil. Trans. Roy. Soc. A*, **237**, 411 (1938).
31a. Booker, H. G., *Jour. Geophys. Res.*, **54**, 243 (1949).
32. Schekulin, *Hochfrequenztech. u. Electroakust.*, **36**, 172 (1936).
32a. Whale, H. A. and Stanley, J. P., *Jour. Atmos. Terr. Phys.*, **1**, 82 (1950).
32b. Poeverlein, H., *Zeits. für. Angewandte Physik*, **2**, 152 (1950).
33. Gamow, G., *Atomic Nuclei and Nuclear Transformation* (Oxford University Press), p. 94 (1937).
34. Saha, M. N. and Rai, R. N., *Proc. Nat. Inst. Sci. (India)*, **3**, 359 (1937).
35. Deb, A. C., *Ind. Jour. Phys.*, **14**, 451 (1940).
36. Kemble, E. C., *Phys. Rev.*, **48**, 549 (1935).
37. Rydbeck, O., *Trans. Chalm. Univ. Gott., Sweden*, No. 74 (1948).
37a. Stanley, J. P., *Canad. Jour. Res. A*, **28**, 549 (1950).
37b. Stanley, J. P., *Jour. Atmos. Terr. Phys.*, **1**, 65 (1950).
38. Wilkes, M. V., *Proc. Roy. Soc. A*, **175**, 143 (1940).
39. Wilkes, M. V., *Proc. Roy. Soc. A*, **189**, 130 (1947).
40. Rawer, M., *Ann. Phys. Leipzig*, **35**, 385 (1939).
41. Bose, S. N., *Ind. Jour. Phys.*, **21**, 121 (1938).
42. Saha, M. N. and Banerjea, B. K., *Ind. Jour. Phys.*, **28**, 159 (1945).
43. Banerjea, B. K., *Proc. Roy. Soc. A*, **190**, 67 (1947).
44. Saha, M. N., Banerjea, B. K. and Guha, U. C., *Ind. Jour. Phys.*, **30**, 181 (1947).
45. Appleton, E. V., *U.R.S.I., Vol. VIII, Mixed Commission on Ionosphere*, p. 9 (1949).
46. Appleton, E. V. and Barnett, M. A. F., *Electrician*, **95**, 678 (1925).
47. Rakshit, H., *Phil. Mag.*, **12**, 897 (1931).
48. Hollingworth, J., *Proc. Roy. Soc. A*, **119**, 444 (1928); *Proc. I.E.E. (Wireless)*, **1**, 57 (1926).
49. Thomas, H. A. and Chalmers, R. G., *Jour. I.E.E., Pt. III*, **95**, 7 (1948).
50. Higgs, A. J., *Coun. Sci. Ind. Res. Org., Australia*, Report No. PD 25/2.
51. Stoffregen, W., *Terr. Mag. and Atmos. Elec.*, **53**, 269 (1948).
52. Lindquist, R., *Trans. Chalm. Univ. No.* **109** (1951).
52a. Wadley, T. L., *Jour. I.E.E. Pt. III*, 96, 483 (1949).
52b. Sulzer, P. G., *Jour. Appl. Phys.*, **20**, 187 (1949).
53. Rydbeck, O. E. H., *Jour. Appl. Phys.*, **21**, 1205 (1950).
53a. Delobeau, F., Harnischmacher, E. and Obreil, F., *Rev. Sci., Paris*, **88**, 17 (1950).
53b. Appleton, E. V. and Builder, G., *Proc. Phys. Soc. (London)*, **45**, 208 (1933).
54. Appleton, E. V., *Nature*, **133**, 793 (1934).

55. Baral, S. S. and Mitra, A. P., *Jour. Atmos. Terr. Phys.*, **1**, 95 (1950).
56. Scott, J. C. W., *Terr. Mag. and Atmos. Elec.*, **53**, 109 (1948).
56a. Whale, H. A., *Jour. Atmos. Terr. Phys.*, **1**, 233 (1951).
57. Menzel, D. H. and Bailey, D. K., *Relations entre les Phénomènes Solaires et Géophysiques, Colloques Internationaux, Lyon* (1947); published by Rev. d'Opt.
58. Mimno, H. R., *Nature*, **134**, 63 (1934).
59. Ratcliffe, J. A. and White, E. L. C., *Proc. Phys. Soc.* (*London*), **45**, 399 (1933); **46**, 107 (1934).
60. Kirby, S. S. and Judson, E. B., *Proc. I.R.E.*, **23**, 733 (1935).
61. Appleton, E. V. and Ratcliffe, J. A., *Proc. Roy. Soc. A*, **128**, 133 (1930).
62. Appleton, E. V., *Int. Union Sci. Rad. Tel.*, **1**, Part 1 (1928).
63. Mitra, S. K. and Syam, P., *Nature*, **135**, 953 (1935).
64. Ellyett, C. D., *Terr. Mag. and Atmos. Elec.*, **53**, 1 (1947).
65. Bracewell, R. N., Budden, K. G., Ratcliffe, J. A., Straker, T. W. and Weekes K., *Proc. I.E.E.*, Part II, **98**, 221 (1951).
66. Bates, D. R. and Seaton, M. J., *Proc. Roy. Soc.*, *B*, **63**, Part II, 129 (1950).
67. Pfister, W., *Jour. Geoph. Res.*, **54**, 315 (1949).
68. Syam, P., *Ind. Jour. Phys.*, **10**, 13 (1936).
69. Martyn, D. F., *Proc. Phys. Soc.* (*London*), **47**, 323 (1935).
70. Hartree, D. R., *Proc. Camb. Phil. Soc.*, **27**, 143 (1930-31).
71. Ginsburg, V., *Jour. Phys. U.S.S.R.*, **8**, 383 (1944).
72. Booker, H. G. and Berkner, L. V., *Nature*, **141**, 562 (1938); *Terr. Mag. and Atmos. Elec.*, **43**, 427 (1938).
73. Martyn, D. F. and Munro, G. H., *Terr. Mag. and Atmos. Elec.*, **44**, 1 (1939).
74. Appleton, E. V., Farmer, F. T. and Ratcliffe, J. A., *Nature*, **141**, 409 (1938).
75. Farmer, F. T. and Ratcliffe, J. A., *Nature*, **135**, 831 (1935).
76. Ratcliffe, J. A., *Proc. Phys. Soc.* (*London*), **51**, 747 (1939).
77. Smith, N., *Bur. Stds. Jour. Res.*, **26**, 105 (1941).
78. Farmer, F. T., Childs, C. B. and Cowie, A., *Proc. Phys. Soc.* (*London*), **50**, 767 (1938).
79. Beynon, W. J. G., *Dept. Sci. and Ind. Res.*, England, *Paper No. R.R.B./C. 82* (1943).
80. Kelso, J. M., *Proceedings of the Conference on Ionospheric Research* (*Pennsylvania*), July, 1949, pp. 16–21.
81. Circular No. 462, *U.S. Department of Commerce, Nat. Bur. Stds., Washington* (1948).
82. Farmer, F. T. and Ratcliffe, J. A., *Proc. Phys. Soc.* (*London*), **48**, 839 (1936).
83. Briggs, B. H., *Proc. Phys. Soc. B*, **64**, 255 (1951).
84. Beynon, W. J. G., *Jour. I.E.E.*, **95**, 325 (1948).
85. Martyn, D. F., *Proc. Phys. Soc.* (*London*), **47**, 323 (1935).
86. Appleton, E. V., Beynon, W. J. G. and Piggott, R., *Nature*, **70**, 695 (1946); Piggott, R., *Electronics*, December (1946).
87. White, F. W. G. and Brown, L. W., *Proc. Roy. Soc. A*, **153**, 639, (1936).
88. White, F. W. G. and Straker, T. W., *Proc. Phys. Soc.* (*London*), **51**, 865 (1939).
89. Best, J. E. and Ratcliffe, J. A., *Proc. Phys. Soc.* (*London*), **50**, 233 (1938).
90. Jaeger, J. C., *Proc. Phys. Soc.* (*London*), **59**, 87 (1947).
91. Hacke, J. E. and Kelso, J. M., *Proc. I.R.E.*, **36**, 1477 (1948); Guha, U. C. *Jour. Geoph. Res.*, **54**, 355 (1949).
92. Rydbeck, O., *Phil. Mag.*, **34**, 251 (1943).
93. Booker, H. G. and Seaton, S. L., *Intern. Assn. of Terr. Mag. and Elec., Washington Assembly*, Sept., 1939.
94. Appleton, E. V. and Beynon, W. J. G., *Proc. Phys. Soc.* (*London*), **52**, 518 (1940),
95. Pierce, J. A., *Phys. Rev.*, **71**, 698 (1947).
96. Fisk, J. B., *Phys. Rev.*, **49**, 167 (1936).
97. Mitra, S. K., Ray, B. B. and Ghosh, S. P., *Nature*, **145**, 1017 (1940).

98. Ray, B. B., *Science and Culture (Calcutta)*, 3, 679 (1937-38).
99. Huxley, L. H. G. and Zaazou, A. A., *Proc. Roy. Soc. A*, 196, 402 (1949).
99a. Huxley, L. H. G., *Proc. Roy. Soc. A*, 196, 427 (1949).
100. Townsend, J. S. and Tizard, H. T., *Proc. Roy. Soc. A*, 88, 347 (1913).
101. Elias, G. J., *Jahr. d. drahtl. Tel.*, 27, 66 (1926).
102. Eckersley, T. L., *Nature*, 135, 435 (1935).
103. Toshniwal, G. R., Pant, B. D. and Bajpai, R. R., *Proc. Nat. Inst. Sci. (India)*, 3, 337 (1937).
104. Appleton, E. V., *Nature*, 135, 618 (1935).
105. Martyn, D. F., Piddington, J. H. and Munro, G. H., *Proc. Roy. Soc. A*, 158, 536 (1937).
106. Appleton, E. V. and Ratcliffe, J. A., *Proc. Roy. Soc. A*, 117, 576 (1928).
107. Ratcliffe, J. A. and White, E. L. C., *Phil. Mag.*, 16, 125 (1933).
108. Eckersley, T. L. and Farmer, F. T., *Proc. Roy. Soc. A*, 184, 196 (1945).
109. Appleton, E. V. and Beynon, W. J. G., *Proc. Phys. Soc.*, 59, 58 (1947).
110. Smith, N., *Bur. Stds. Jour. Res.*, 19, 89 (1937); *Proc. I.R.E.*, 27, 332 (1939).
111. Millington, G., *Proc. Phys. Soc. (London)*; 50, 801 (1938).
112. Naismith, R., *Dept. Sci. and Ind. Res.*, England, *Paper No. R.R.B./C. 75*, (1942).
113. Dellinger, J. H. and Smith, N., *Proc. I.R.E.*, 36, 258 (1948).
113a. Bibl, K., *Rev. Sci., Paris*, 88, 27 (1950).
113b. Benner, A. H., *Proc. I.R.E.*, 37, 44 (1949).
114. Chapman, S., *Proc. Phys. Soc. (London)*, 43, 26 and 433 (1931).
115. Pannekoek, A., *Proc. Amsterd. Acad.*, 29, 1165 (1926).
116. Saha, M. N., *Phil. Mag.*, 1, 1025 (1920).
117. Woltjer, J., *Physica*, 5, 406 (1925).
118. Milne, E. A., *Phil. Mag.*, 47, 209 (1924).
119. Bhar, J. N., *Ind. Jour. Phys.*, 12, 363 (1938).
120. Appleton, E. V., *Proc. Roy. Soc. A*, 162, 451 (1937); *Quar. Jour. Roy. Met. Soc.*, 65, 324 (1939).
121. Nicolet, M. and Bossy, L., *Ann. de Géoph.*, 5, 275 (1949).
121a. Gledhill, J. A. and Szendrei, M. E., *Proc. Phys. Soc. B*, 63, 427 (1950).
122. Saha, M. N. and Rai, R. N., *Proc. Nat. Inst. Sci. (India)*, 4, 319 (1938).
123. Mitra, S. K., *Nature*, 142, 914 (1938).
124. Wulf, O. R. and Deming, L. S., *Terr. Mag. and Atmos. Elec.*, 43, 283 (1938).
125. Nicolet, M., *Mem. Roy. Met. Inst. Bel.*, 19, 1 (1945).
126. Mohler, F. L., *Bur. Stds. Jour. Res.*, 25, 507 (1940).
126a. Bradbury, N. E., *Terr. Mag. and Atmos. Elec.*, 43, 55 (1938).
127. Bates, D. R. and Massey, H. S. W., *Proc. Roy. Soc. A*, 187, 261 (1946).
128. Hoyle, F. and Bates, D. R., *Terr. Mag. and Atmos. Elec.*, 53, 51 (1948).
129. Bates, D. R., *Proc. Roy. Soc. A*, 196, 562 (1949).
129a. Mitra, S. K., Bhar, J. N. and Ghosh, B. N., *Ind. Jour. Phys.*, 12, 455 (1938).
130. Pfister, W., *Proceedings of the Conference on Ionospheric Research*, K1–K15 (1949).
131. Vassy, A. and Vassy, E., *Cahiers de Physique*, 9, 28 (1942).
132. Gauzit, J., *C.R. Acad. Sci. (Paris)*, 217, 179 (1943).
133. Appleton, E. V. and Naismith, R., *Proc. Phys. Soc., B*, 60, 59 (1947).
134. Martyn, D. F., *Proc. Roy. Soc. A*, 194, 445 (1948).
135. Wooley, R. v.d.R., *Proc. Roy. Soc. A*, 187, 403 (1946).
136. Massey, H. S. W., *Paper presented at Lyons Conference*, 1947.
137. Waldmeier, M., *Terr. Mag. and Atmos. Elec.*, 52, 333 (1947).
137a. Appleton, E. V. and Chapman, S., *Proc. I.R.E.*, 23, 658 (1935).
137b. Baral, S. S. and Mitra, S. N., *Science and Culture (Calcutta)*, 10, 175 (1944-45).
137c. Kirby, S. S., Berkner, L. V., Gilliland, T. R. and Norton, K. A., *Bur. Stds. Jour. Res.*, 11, 829 (1933).
138. Oisawa, K., *Report of Ionospheric Research in Japan*, 4, No. 1, 53 (1950).

139. Gejer, S., Akerlind, P., *Terr. Mag. and Atmos. Elec.*, **52**, 479 (1947).
140. Vyeda, H., Kudo, H., Shimiju, J. and Sato, R., *Report of Ionospheric Research in Japan*, **4**, No. 2, 75 (1950).
141. Nagata, Y., *Report of Ionospheric Research in Japan*, **4**, No. 1, 21 and 54 (1950).
141*a*. Smithrose, R. L., *J. Brit. Inst. Rad. Eng.*, **6**, 82 (1946).
141*b*. Schafer, J. P. and Goodall, W. M., *Science*, **76**, 444 (1932).
141*c*. Denisse, J. F., Seligmann, P. and Gallet, R., *C.R. Acad. Sci., Paris*, **225**, 1169 (1947).
142. Kiepenheur, K. O., *Month. Not. Roy. Astr. Soc.*, **106**, 515 (1946).
143. Osawa, K., *Report of Ionospheric Research in Japan*, **4**, No. 2, 125 (1950).
144. Wooley, R.v.d.R., U.R.S.I., Vol. VIII, *Mixed Commission on ionosphere*, p. 85 (1949).
145. Danisse, J. F., Seligmann, P. and Gallet, R., *C.R. Acad. Sci., Paris*, **225**, 1169 (1947).
146. Budden, K. G., Ratcliffe, J. A. and Wilkes, M. V., *Proc. Roy. Soc. A*, **171**, 188 (1939).
147. Benner, A. H., *Proc. I.R.E.*, **38**, 685 (1950); *ibid.*, **39**, 186 (1951).
148. Berkner, L. V. and Wells, H. W., *Terr. Mag. and Atmos. Elec.*, **43**, 15 (1938); **41**, 173 (1936); **44**, 313 (1939).
149. Appleton, E. V. and Piggot, R., *Nature*, **165**, 130 (1950).
150. Appleton, E. V. and Naismith, R., *Proc. Roy. Soc. A*, **150**, 685 (1935).
151. Mathur, L. S., *Proc. Nat. Inst. Sci. (India)*, **7**, 222 (1937).
152. Martyn, D. F. and Pulley, O. O., *Proc. Roy. Soc. A*, **154**, 455 (1936).
153. Wells, H. W., *Jour. Geophys. Res.*, **54**, 277 (1949).
153*a*. Appleton, E. V., *Jour. Atmos. Terr. Phys.*, **1**, 106 (1950).
154. Appleton, E. V., *Nature*, **157**, 691 (1946).
155. Liang, P. H., *Nature*, **160**, 642 (1947).
156. Bailey, D. K., *Terr. Mag. and Atmos. Elec.*, **53**, 35 (1948).
157. Lung, H. L., *Jour. Geophys. Res.*, **54**, 177 (1949).
157*a*. Gallet, R. and Rawer, K., *Ann. Géophys.*, **6**, 104, 212 (1950).
158. Mitra, S. K., *Nature*, **158**, 668 (1946).
159. Appleton, E. V. and Naismith, R., *Phil. Mag.*, **27**, 147 (1939).
160. Smith, N., Gilliland, T. R. and Kirby, S. S., *Bur. Stds. Jour. Res.*, **21**, 835, 1159 (1938).
161. Waldemier, M., *Hebvetia Physica Acta*, **17**, 168 (1944).
162. Wells, H. W., *Terr. Mag. and Atmos. Elec.*, **52**, 315 (1947).
163. Allen, C. W., *Terr. Mag. and Atmos. Elec.*, **53**, 433 (1948).
164. Bartels, J., *Jour. Atmos. Terr. Phys.*, **1**, 2 (1950).
165. Massey, H. S. W. and Bates, D. R., *Rep. Prog. Phys.*, Physical Society, London, **9**, 62 (1942-43).
165*a*. Gaydon, A. G., *Nature*, **153**, 407 (1944).
166. Bates, D. R., *Phys. Rev.*, **78**, 492 (1950).
166*a*. Bates, D. R., Buckingham, R. A., Massey, H. S. W. and Unwin, J. J., *Proc. Roy. Soc. A*, **170**, 322 (1939).
166*b*. Mitra, S. K., *Nature*, **167**, 897 (1951).
167. Bates, D. R. and Massey, H. S. W., *Phil. Trans. Roy. Soc. A*, **239**, 269 (1943).
168. Lozier, W. W., *Phys. Rev.*, **46**, 268 (1934).
169. Vier, D. T. and Mayer, F. E., *Jour. Chem. Phys.*, **12**, 28 (1944).
170. Massey, H. S. W., *Negative Ions, Cambridge University Press*, p. 74 (1938).
171. Tukada, T., *Rep. Rad. Res. (Japan)*, **7**, 121 (1937).
172. Massey, H. S. W. and Smith, R. A., *Proc. Roy. Soc. A*, **155**, 472 (1936).
173. Mitra, S. K., *Science and Culture (Calcutta)*, **9**, 46 (1943-44).
174. Bates, D. R. and Seaton, M. J., *Proc. Phys. Soc. B*, **63**, 129 (1950).
175. Savitt, J., *Jour. Geoph. Res.*, **55**, 385 (1950).
176. Hulbert, E. H., *Phys. Rev.*, **55**, 870 (1939).

177. Berkner, L. V. and Seaton, S. L., *Terr. Mag. and Atmos. Elec.*, **45**, 393 (1940).
178. Seaton, S. L., *Jour. Met.*, **5**, 5 (1948).
179. Mitra, A. P., *Jour. Atmos. Terr. Phys.*, **1**, 286 (1951).
180. Bates, D. R., *Mon. Not. Roy. Astr. Soc.*, **109**, 216 (1949).
181. Seaton, S. L., *Phys. Rev.*, **72**, 712 (1947).
182. Berkner, L. V., '*Terrestrial Magnetism and Electricity*' (Edited by J. A. Fleming) McGraw Hill Book Co. (1939).
183. Bates, D. R. and Massey, H. S. W., *Proc. Roy. Soc.*, **187**, 261 (1946).
184. Sayers, J. A., *Rep. Prog. Phys.*, **9**, 52 (1942-43).
185. Ghosh, S. N., *Proc. Nat. Inst. Sci., India*, **10**, 333 (1944).
186. Yonezawa, T., *Report of Ionospheric Research in Japan*, **4**, 79 (1950).
187. Biondi, M. A. and Brown, S. C., *Phys. Rev.*, **75**, 1700 (1949); **76**, 1697 (1949).
188. Holt, R. B., Richardson, J. M., Howland, B., and Mc Clure, B. T., *Phys. Rev.*, **77**, 239 (1950).
189. Contract No. AF 19(122)–89, *Geophysical Research Directorate, A.M.C., U.S.A.F.*, *Quarterly Progress Report*, Nos. 1, 2 and 3.
190. Ghosh, S. P., *M.Sc. Thesis* (*Unpublished*), *Calcutta University*, 1938.
191. Penndorf, R., *Jour. Met.*, **5**, 152 (1948).
192. *Interim Progress Report, Electronics Research Laboratory*, Stanford University, California, Bur. Stds. Contract No. Cst.–10751 (1951).
193. Appleton, E. V., and Naismith, R., *Proc. Roy. Soc. A*, **150**, 685 (1935).
194. Wells, H. W., *Jour. Geoph. Res.*, **54**, 277 (1949).
195. Tremmellen, K. W. and Cox, J. W., *Jour. I.E.E.*, **94**, Part IIIA, 200 (1947).
196. Appleton, E. V. and Naismith R., *Proc. Phys. Soc.*, **52**, 402 (1940).
197. Skellett, A. M., *Proc. I.R.E.*, **20**, 1933 (1932).
198. Schafer, J. P. and Goodall, W. M., *Proc. I.R.E.*, **20**, 1941 (1932).
199. Mitra, S. K., Syam, P., and Ghosh, B. N., *Nature*, **133**, 533 (1933).
200. Bhar, J. N., *Ind. Jour. Phys.*, **12**, 109 (1937).
201. Pierce, J. A., *Phys. Rev.*, **71**, 88 (1947).
202. Pineo, V. C., *Science*, **110**, 280 (1949); *Ibid.*, **112**, 50 (1950).
203. Ranzi, I., *Nature*, **130**, 368 (1932).
204. Martyn, D. F. and Pulley, O. O., *Proc. Roy. Soc. A*, **154**, 455 (1936).
205. Best, J. E., Farmer, F. T. and Ratcliffe, J. A., *Proc. Roy. Soc. A*, **164**, 96 (1938).
206. Berkner, L. V. and Wells, H. W., *Terr. Mag. and Atmos. Elec.*, **42**, 183 and 301 (1937).
207. Bhar, J. N. and Syam, P., *Phil. Mag.*, **23**, 513 (1937).
208. Appleton, E. V. and Naismith, R., *Proc. Phys. Soc.* (*London*), **45**, 389 (1933).
209. Wilson, C. T. R., *Proc. Phys. Soc.* (*London*), **37**, 320 (1925).
210. Schonland, B. F. J., *Atmospheric Electricity* (*Methuen*), (1932).
211. Appleton, E. V., Watson Watt, R. A. and Herd, J. F., *Proc. Roy. Soc. A*, **111**, 615 (1926).
212. Healey, R. H., *Tech. Rev.*, **3**, 215 (1938).
213. Mögel, H., *Wireless Engineer*, **8**, 604 (1931); *E.N.T.*, **8**, 321 (1931).
214. Eckerseley, T. L., *Jour. I.E.E.*, **86**, 548 (1940).
215. Silberstein, R., *National Bureau of Standards* (*Research Paper No. Rp 2071*), **44**, 199 (1950).
216. Dieminger, W., 'Origin of ionospheric scattering,' *Paper presented at the Conference on Ionospheric Research, State College*, Pennsylvania, U.S.A. (1950).
217. Peterson, M., 'Interpretation of Long-scatter Echo Pattern,' *Paper presented at the Conference on Ionospheric Research, State College*, Pennsylvania, U.S.A., 1949.
218. Rivault, R., *Proc. Phys. Soc. B*, **63**, 126 (1950).

219. Findlay, J. W., *Proc. Phys. Soc. B*, **63**; 148 (1950); Ferrell, O. P., *Bull. Amer. Met. Soc.*, **25**, 371 (1944); *C.Q. February*, p. 25 (1949).
220. Lindquest, R., 'Ionospheric Effects of Solar Flares,' *Report from the Research Laboratory of Electronics, Chalmers University of Technology*, Gothenburg, Sweden, 1950.
221. Ratcliffe, J. A., *Nature*, **162**, 9 (1948).
222. Mitra, S. N., *Jour. I.E.E.*, Pt. III, **96**, 505 (1949).
223. Khastgir, S. R. and Das, P. M., *Ind. Jour. Phys.*, **24**, 277 (1950).
224. Sen Gupta, M. M. and Dutt, S. K., *Ind. Jour. Phys.*, **15**, 447 (1941).
224a. Banerji, R. B., *Ind. Jour. Phys.*, **24**, 359 (1951).
225. Briggs, B. H., Phillips, G. J. and Shinn, D. H., *Proc. Phys. Soc. B*, **63**, 106 (1950).
226. Gerson, N. C., *Nature*, **166**, 316 (1950).
227. Blackett, P. M. S. and Lovell, A. C. B., *Proc. Roy. Soc. A*, **177**, 183 (1941).
228. Booker, H. G. and Gordon, W. E., *Proc. I.R.E.*, **38**, 401 (1950).
229. Booker, H. G., 'The Theory of the scattering of radio waves in the troposphere and ionosphere,' *Paper read before the Symposium on the Theory of Electromagnetic Waves held at New York University*, 1950.
230. Rawer, K., *Nature*, **163**, 529 (1949).
231. Dellinger, J. H., *Science*, **82**, 351 (1935).
232. McNish, A. G., *Terr. Mag. and Atmos. Elec.*, **42**, 109 (1937).
232a. Bracewell, R. N. and Straker, T. W., *Mon. Not. Roy. Astr. Soc.*, **109**, 28 (1949).
233. Budden, K. G., Ratcliffe, J. A. and Wilkes, M. V., *Proc. Roy. Soc. A*, **171**, 188 (1939).
234. Bureau, R., *C.R. Acad. Sci. (Paris)*, **203**, 1257 (1936); *Proc. Phys. Soc. B*, **63**, 122 (1950).
235. Weekes, K., *Nature*, **165**, 935 (1950).
236. Williams, C., *Proc. Phys. Soc., B*, **63**, 143 (1950).
237. Bracewell, R. N., *Proc. Phys. Soc., B*, **63**, 143 (1950).
238. Smith-Rose, R. L., *Nature*, **165**, 37 (1950).
239. Findlay, J. W., *Nature*, **159**, 58 (1947).
240. Martyn, D. F., Munro, G. H., Higgs, A. H. and Williams, A. C., *Nature*, **140**, 603 (1937).
241. Becker, W. and Dieminger, W., *Zeit. f. Naturforschung, Band* **5**, Haft **6**, p. 308 (1950).
242. Waldmeier, M., *Z. für. Astrophys.*, **20**, 46 (1940).
243. Ellison, M. A., *Mon. Not. Roy. Astr. Soc.*, **103**, 3 (1943).
244. Jones, M. W., *Trans. Amer. Geophys. Union*, **31**, 187 (1950).
245. Wells, H. W., *Terr. Mag. and Atmos. Elec.*, **52**, 315 (1947).
246. Harang, L., *Terr. Mag. and Atmos. Elec.*, **42**, 55 (1937); *Ibid.*, **43**, 41 (1938).
247. Appleton, E. V., Naismith, R. L. and Ingram, L. J., *Phil. Trans. Roy. Soc. A*, **236**, 191 (1937).
248. Martyn, D. F., *Proc. Roy. Soc. A*, **189**, 241 (1947).
249. Matsushita, S., *Report of Ionospheric Research in Japan*, **4**, 47 (1950).
250. Störmer, C., Night Clouds Observed in Norway, *Vid. Akad. Arch.* I.M.—N. Kl. No. 2 (1933).
251. Jesse, O., *Sitz. der Kgl. Preuss. Akad. der Wissenschafter* (1890, 1891).
251a. Manning, L. A., Villard, O. G., Jr., Peterson, A. M., *Tech. Rep.*, **22**; *Electronics Research Laboratory, Stanford University* (1950); *Proc. I.R.E.* (In press); Paton, J., *Proc. Phys. Soc. B*, **63**, 1039 (1950).
251b. Hey, J. S., *Nature*, **159**, 119 (1947).
252. Mimno, H. R., *Rev. Mod. Phys.*, **9**, 1 (1937).
252a. Meek, J. H., *Jour. Geophys. Res.*, **54**, 284 (1949).
252b. Mitra, S. N., *Jour. I.E.E., Pt. III*, **96**, 441 (1949).

252c. Aono, T. and Minoguchi, T., *Report of Ionospheric Research in Japan*, **4**, No. 2, p. 103 (1950).
253. Pawsey, J. L., *Proc. Camb. Phil. Soc.*, **31**, 125 (1935).
253a. Munro, G. H., *Nature*, **162**, 886 (1948).
254. Beynon, W. J. G., *Nature*, **162**, 887 (1948).
255. McNicol, R. W. E., *Jour. I.E.E., Pt. III*, **96**, 517 (1949).
256. Burkard, G., *Terr. Mag. and Atmos. Elec.*, **53**, 61 (1948).
257. Martyn, D. F., *Proc. Roy. Soc. A*, **190**, 273 (1947).
258. Martyn, D. F., *Proc. Roy. Soc. A*, **194**, 429 and 445 (1948).
259. Martyn, D. F., *Nature*, **163**, 34 (1949).
260. Martyn, D. F., *Report on the Tidal Oscillations in the Ionosphere* (In course of publication).
260a. Mitra, A. P., *Ind. Jour. Phys.*, **24**, 387 (1950).
261. Appleton, E. V. and Weekes, K., *Proc. Roy. Soc. A*, **171**, 171 (1939).
262. Appleton, E. V. and Beynon, W. J. G., *Nature*, **164**, 308 (1949).
263. Jones, M. W. and Jones, J. G., *Jour. Met.*, **7**, 14 (1950).
264. Chowdhury, D. C., *Ind. Jour. Phys.*, **25**, 1 (1951).
265. Kirkpatrick, C. B., *Aust. Jour. Sci. Res. A*, **1**, 423 (1948).
266. Mitra, A. P., *Jour. Terr. Atmos. Phys.* **1**, 286 (1951).
267. Pierce, J. A. and Mimno, H., *Phys. Rev.*, **57**, 95 (1940).
268. Munro, G. H., *Nature*, **162**, 886 (1948); *Ibid.*, **163**, 813 (1949).
269. Martyn, D. F., *Proc. Roy. Soc.*, **201**, 216 (1950).
270. Wells, H. W., Watts, J. M. and George, D. E., *Phys. Rev.*, **69**, 540 (1946).
271. Ross, W. and Bramley, E. N., *Nature*, **164**, 355 (1949).
272. Ross, W. and Bramley, E. N., *Nature*, **159**, 132 (1947).
273. Ross, W., Department of Scientific and Industrial Research, *Radio Research Special Report No. 19*, England (1949).
274. Bailey, V. A., and Martyn, D. F., *Nature*, **133**, 218 (1934); *Phil. Mag.*, **18**, 369 (1934); *W.E. and E.W.*, **12**, 122 (1935).
275. Tellegen, B. D. H., *Nature*, **131**, 840 (1933).
276. Van der Pol, B., and Van der Mark, I., *U.R.S.I. Report*, September (1934).
277. Bailey, V. A., *Nature*, **139**, 68 (1937); *Phil. Mag.*, **26**, 425 (1938); *Nature*, **142**, 613 (1938).
278. Bailey, V. A., *Wireless World*, **40**, 204 (1937).
279. Townsend, J. S. and Tizard, H. T., *Proc. Roy. Soc. A*, **88**, 347 (1913).
280. Huxley, L. G. H. and Ratcliffe, J. A., *Proc. I.E.E., Pt. III*, **96**, 443 (1949), summary of the work done till 1949.
281. Bailey, V. A., *Phil. Mag.*, **23**, 929 (1937).
282. Huxley, L. G. H., Foster, H. G. and Newton, C. C., *Nature*, **159**, 300 (1947).
283. Ratcliffe, J. A., and Shaw, I. J., *Proc. Roy. Soc. A*, **193**, 311 (1948).
284. Huxley, L. G. H., Foster, H. G. and Newton, C. C., *Proc. Roy. Soc. A*, **61**, 134 (1948).
285. Bailey, V. A., *Phil. Mag.*, **23**, 774 (1937).
286. Bailey, V. A., *Nature*, **139**, 838 (1937).
287. Cutolo, M., Corlevaro, M. and Gherghi, M., *Alta Frequenza*, **15**, 111 (1946).
288. Cutolo, M., *Nature*, **160**, 834 (1947).
289. Cutolo, M., *Nature*, **167**, 315 (1951).
290. Shaw, I. J., *Proc. Phys. Soc. B*, **63**, 1 (1951).
291. Gherzi, E., *Nature*, **165**, 38 (1950).
292. Martyn, D. F. and Pulley, O. O., *Proc. Roy. Soc. A*, **154**, 455 (1936); Gerson, N. C., *Proc. I.R.E.*, **38**, 1456 (1950).
293. Martyn, D. F. and Pulley, O. O., *Proc. Roy. Soc. A*, **154**, 455 (1936).
294. Bannon, J., Higgs, A. J., Martyn, D. F. and Munro, G. H., *Proc. Roy. Soc. A*, **174**, 298 (1940).

CHAPTER VII

ELECTRICAL CURRENTS IN THE UPPER ATMOSPHERE: TERRESTRIAL MAGNETIC VARIATIONS

1. Vestine, E. H., Laporte, L., Lance, I. and Scott, W. E., *The Geomagnetic Field, Its Description and Analysis*, Publication No. 580, p. 4, Carnegie Institution, Washington, D.C. (1947).
2. Blackett, P. M. S., *Nature*, **159**, 658 (1947).
2a. Schuster, A., *Proc. Phys. Soc., London*, **24**, 121 (1912).
3. Sutherland, T., *Terr. Mag. and Atmos. Elec.*, **13**, 155 (1908).
4. Babcock, H. W., *Astrophys. Jour.*, **105**, 105 (1947).
5. Thiessen, G., *Zeit. Astrophys.*, **26**, 16 (1949).
6. Babcock, H. W., *Publ. Astr. Soc. Pacif.*, **61**, 226 (1949).
7. Bullard, E. C., *Proc. Roy. Soc. A*, **197**, 433 (1949).
8. Elsasser, W. M., *Phys. Rev.*, **69**, 106 (1946); **70**, 202 (1946); **72**, 821 (1947).
9. Bullard, E. C., *Mon. Not. Roy. Astr. Soc., Geophys. Suppl.*, **5**, 248 (1948).
9a. Bartels, J., *Int. Ass. Terr. Mag. Elec.*, Bull. No. 12e, Washington (1951).
10. Bartels, J., *Terr. Mag. and Atmos. Elec.*, **37**, 1 (1932); **39**, 1 (1934); **40**, 265 (1945); **41**, 374 (1936); **43**, 131 (1938).
10a. Bartels, J., Heck, N. H. and Johnston, H. F., *Terr. Mag. and Atmos. Elec.*, **44**, 441 (1939).
11. Chapman, S. and Miller, J. C. P., *Mon. Not. Roy. Astr. Soc., Geophys. Suppl.*, **4**, 649 (1940).
12. Chapman, S., *Proc. Roy. Soc. A*, **115**, 242 (1927).
13. Chapman, S., *Proc. Roy. Soc. A*, **95**, 61 (1919).
14. Chapman, S., *Phil. Trans. Roy. Soc. A*, **225**, 49 (1925).
15. Hasegawa, M., *Proc. Imp. Acad. Tokyo*, **12**, 88, 185, 221, 225, 277 (1936).
16. Gauss, C. F., 'Allgemeine Theorie des Erdmagnetismus,' in *Resultate Magn. Verein* 1838, reprinted in *Werke*, Band **5**, p. 121.
17. Schuster, A., *Phil. Trans. Roy. Soc. A*, **180**, 467 (1889); **208**, 163 (1908).
18. Chapman, S., *The Earth's Magnetism*, Mathuen & Co. Ltd., London (1936).
19. Chapman, S. and Whitehead, T. T., *Trans. Phil. Soc. Camb.*, **22**, 463 (1923).
20. Bartels, J., *Handbuch der Expt. Phys.*, **25**, 640 (1928).
21. Chapman, S., *Phil. Trans. Roy. Soc. A*, **218**, 1 (1919).
22. Terada, K., *Terr. Mag. and Atmos. Elec.*, **52**, 189 (1947).
22a. Martyn, D. F., *Nature*, **163**, 34 (1949).
23. Bartels, J. and Johnston, H. F., *Terr. Mag. and Atmos. Elec.*, **45**, 269, 485 (1950).
24. Balfour Stewart, *Ency. Brit.* (Ninth Edition), p. 36 (1882).
25. Gunn, R., *Phys. Rev.*, **32**, 153 (1928); *Terr. Mag. and Atmos. Elec.*, **34**, 7 (1929).
26. Chapman, S., *Proc. Roy. Soc. A*, **122**, 369 (1929).
27. Pedersen, P. O., *The Propagation of Radio Waves*, published by 'Danmarks Naturvidenskabelige Samfund', Copenhagen (1927).
28. McNish, A. G., *Terrestrial Magnetism and Electricity*, Vol. VIII of series 'Physics of the Earth', McGraw-Hill, p. 345 (1939).
29. Egedal, J., *Terr. Mag. and Atmos. Elec.*, **52**, 449 (1947).
29a. Vestine, E. H., Laporte, L., Lance, I. and Scott, W. E., *The Geomagnetic Field, Its Description and Analysis*, Publication No. 580, p. 359, Carnegie Institution, Washington, D.C. (1947).
30. Slaucitajs, L. and McNish, A. G., *Trans. Edinburgh Meeting* (1936); *Internat. Union Geod. and Geophys., Assoc. Terr. Mag. Elec. Bull.*, No. 10, p. 289 (1937).
30a. Chapman, S., *Proc. Roy. Soc. A*, **115**, 242 (1927).
31. Chapman, S., I.C.S.U., *Mixed Commission on Ionosphere, Proceedings of the First Meeting*, 1948; p. 58, Bruxelles (1949).
32. Chapman, S., *Mon. Not. Roy. Astr. Soc.*, **79**, 70 (1918).

33. Chapman, S., *Geomagnetism*, Vol. II, p. 887, Oxford University Press (1940).
34. Nikolsky, A. P., *Terr. Mag. and Atmos. Elec.*, **52**, 147 (1947).
35. Birkeland, K., *The Norwegian Aurora Polaris Expedition*, 1902-3, Vol. 1, Christiania, H. Aschehong & Co. (1908).
36. Chapman, S., *Terr. Mag. and Atmos. Elec.*, **40**, 349 (1935).
37. Sucksdorff, E., *Terr. Mag. and Atmos. Elec.*, **52**, 201 (1947).
38. Wenner, F., *Bull. U.S. Bur. Stds.*, **12**, 469 (1916).
39. Gish, O. H. and Rooney, W. J., *Terr. Mag. and Atmos. Elec.*, **30**, 161 (1925).
40. Whitehead, S. and Radley, W. G., *Proc. Phys. Soc. (London)*, **47**, 589 (1935).
41. Banerji, S. K., *Memoirs of the India Meteorological Dept.*, **25**, Part I, 1 (1933).
42. Rooney, W. J., *Terrestrial Magnetism and Electricity*, Vol. VIII of series 'Physics of the Earth', McGraw-Hill, p. 270 (1939).
43. Gish, O. H., *Terr. Mag. and Atmos. Elec.*, **28**, 89 (1923).
44. Gish, O. H. and Rooney, W. J., *Terr. Mag. and Atmos. Elec.*, **33**, 79 (1928).
45. Gish, O. H. and Rooney, W. J., *Terr. Mag. and Atmos. Elec.*, **35**, 213 (1930).
46. Rooney, W. J. and Gish, O. H., *Terr. Mag. and Atmos. Elec.*, **32**, 49 (1927).
47. Rooney, W. J. and Gish, O. H., *Terr. Mag. and Atmos. Elec.*, **35**, 61 (1930).
48. Bartels, J., *Terr. Mag. and Atmos. Elec.*, **37**, 1 (1932); **39**, 1 (1934); **40**, 265 (1935); **41**, 374 (1936); **43**, 131 (1938).
49. Chree, C., *Phil. Trans. Roy. Soc. A*, **212**, 75 (1912); **213**, 245 (1913).
50. Ellis, W., *Mon. Not. Roy. Astr. Soc.*, **60**, 142 (1900); **61**, 538 (1901).
51. Maunder, E. W., *Mon. Not. Roy. Astr. Soc.*, **64**, 205 (1904); **65**, 2, 538, 666 (1905); **76**, 63 (1916).
52. Chree, C., *Proc. Roy. Soc. A*, **101**, 368 (1922).
53. Bartels, J., *Terr. Mag. and Atmos. Elec.*, **39**, 201 (1934).
54. Bartels, J., *Elektrische Nachrichten-Technik*, **10** (Sonderhaft), 1 (1933).
55. Chree, C., *Studies in Terrestrial Magnetism*, London, Macmillan & Co. (1912); *Ency. Brit.* (Eleventh Edition), **17**, 353 (1911).
56. Silsbee, H. C. and Vestine, E. H., *Terr. Mag. and Atmos. Elec.*, **47**, 195 (1942).
57. Steiner, L., *Terr. Mag. and Atmos. Elec.*, **26**, 1 (1921).
58. Wiechert Erwin, *Untersuchungen an Baystörungen*, Mitt. Geophys. Warte Königsberg, No. **22**, 26 (1934).
59. Lubiger, F., *Dissertation Univ. Göttingen* (1924).
60. McNish, A. G., *Trans. Edinburg Meeting* (1936); *Internat. Union Geod. and Geophys., Ass. Terr. Mag. Elec. Bull.*, No. 10, p. 282 (1937); *Trans. Am. Geophys. Union*, Seventeenth Annual Meeting, **1**, 166 (1936).
61. Birkeland, K., *Norwegian Aurora Polaris Expedition*, 1902-3, Christiania, **1**, Part 1, 39 (1908); Part 2, 319 (1913).
62. McNish, A. G., *Terr. Mag. and Atmos. Elec.*, **43**, 67 (1938).
63. Wells, H. W., *Terr. Mag. and Atmos. Elec.*, **52**, 315 (1947).
64. La Cour, D., *Terr. Mag. and Atmos. Elec.*, **47**, 265 (1942).
65. Harang, L., *Terr. Mag. and Atmos. Elec.*, **37**, 57 (1932).
66. Rolf, B., *Terr. Mag. and Atmos. Elec.*, **36**, 7 (1931).
67. Eschenhagen, M., *Sitz.-Ber. Akad. Wiss. Berlin*, 678 (1897); *Terr. Mag. and Atmos. Elec.*, **2**, 105 (1897).
68. Angenheister, G., *Göttinger Nachr., Math.-Physik. Kl.*, 565 (1913); *Terr. Mag. and Atmos. Elec.*, **25**, 26 (1920).

CHAPTER VIII

Aurora Polaris

1. Lovell, A. C. B., Clegg, J. A. and Ellyett, C. D., *Nature*, **160**, 372 (1947).
2. Forsyth, P. A., Petrie, W., Vawter, F. and Currie, B. W., *Nature*, **165**, 561 (1950); Davidson, D., *Nature*, **167**, 277 (1951).

3. Fritz, H., *Das Polarlicht* (Leipzig, Brockhaus, 1881).
4. Vestine, E. H., *Terr. Mag. and Atmos. Elec.*, **40**, 77 (1944); see also: *The Geomagnetic Field, Its Description and Analysis* (p. 311, Fig. 149) by Vestine, E. H., Laporte, L., Lance, I. and Scott, W. E., Publication No. 580, Carnegie Institution, Washington, D.C. (1947).
5. White, F. W. G. and Geddes, M., *Terr. Mag. and Atmos. Elec.*, **44**, 367 (1939).
6. Thomson, J. L., *Terr. Mag. and Atmos. Elec.*, **52**, 469 (1947).
7. Störmer, C., *Geofys. Pub.*, **4**, 104 (1924).
8. Vegard, L. and Krogness, O., *Geofys. Pub.*, **1**, No. 1 (1920).
9. Harang, L. and Bauer, W., *Ger. Beitr. z. Geophys.*, **37**, 109 (1932).
10. Simpson, G. C., *Quar. Jour. Roy. Met. Soc.*, **59**, 185 (1933).
11. Störmer, C., *Ger. Beitr. z. Geophys.*, **17**, 254 (1927).
12. Vegard, L., *Geofys. Pub.*, **1**, No. 1, 172 (1920).
13. Störmer, C., *Vid. Selsk. Skr.*, No. 17 (1911).
14. Harang, L. and Tönsberg, E., *Geofys. Pub.*, **9**, No. 5 (1932).
15. See Störmer, C., *Geofys. Pub.*, **12**, No. 7, 3 (1936).
16. Chapman, S., *Terr. Mag. and Atmos. Elec.*, **39**, 299 (1934).
17. Vegard, L. and Krogness, O., *Geofys. Pub.*, **1**, No. 1 (1920).
18. Harang, L. and Tönsberg, E., *Geofys. Pub.*, **9**, No. 5 (1932).
19. McLennan, J. C., Wynne-Edwards, H. S., and Ireton, H. J. C., *Can. Jour. Res.*, **5**, 285 (1931); Currie, B. W., *Terr. Mag. and Atmos. Elec.*, **39**, 293 (1934).
20. Störmer, C., *Geofys. Pub.*, **11**, No. 12, 3 (1936).
21. Vegard, L., *Phil. Mag.*, **42**, 47 (1921).
22. Harang, L., *Terr. Mag. and Atmos. Elec.*, **37**, 167 (1932).
23. Currie, B. W. and Edwards, H. W., *Terr. Mag. and Atmos. Elec.*, **39**, 293 (1934).
24. Vegard, L., *Phil. Mag.*, **23**, 211 (1912).
25. Carlheim-Gyllenskiöld, *Expt. Int. polaris, 1882-83; Expt. Suedois: II Aurorae boreal*, Stockholm, 1886.
26. Boller, W., *Ger. Beitr. z. Geophys.*, **3**, 56 and 550 (1898).
27. Rowland, J. P., *Nature*, **145**, 625 (1940).
28. Stetson, H. T., *Popular Astronomy*, **50**, No. 4 (1942).
29. Dauvillier, M. A., *Jour. de Phys.*, **5**, 398 (1934); *Electrician*, **112**, 787 (1934).
30. Currie, B. W. and Edwards, H. W., *Terr. Mag. and Atmos. Elec.*, **41**, 265 (1936).
31. Rooney, W. J., *Terr. Mag. and Atmos. Elec.*, **39**, 103 (1934).
32. Störmer, C., *Nature*, **123**, 868 (1929).
33. Rayleigh, Lord, *Proc. Roy. Soc. A*, **129**, 458 (1930).
34. Dauvillier, M. A., *Jour. de Phys.*, **5**, 398 (1934).
35. Störmer, C., *Nature*, **140**, 1095 (1937).
36. Vegard, L. and Kvifte, G., *Geofys. Pub.*, **16**, No. 7 (1945).
36*a*. Vegard, L., *Nature*, **165**, 1013 (1950).
36*b*. Nicolet, M. and Dogniaux, R., *Jour. Geophys. Res.*, **55**, 21 (1950).
36*c*. Meinel, A. B., *Astrophys. Jour.*, **112**, 562 (1950).
37. Bernard, R., 'Emission Spectra of the Night Sky and Aurorae', *Gassiot Committee Report* (1947), p. 93, published by the Physical Society, London (1948).
38. Nicolet, M., 'Emission Spectra of the Night Sky and Aurorae', *Gassiot Committee Report* (1947), p. 105, published by the Physical Society, London (1948).
38*a*. Tcheng Mao-Lin and Dufay, J., *Cahiers de Physique*, No. 8, p. 51 (1942); Bernard, R., 'Emission Spectra of the Night Sky and Aurorae', *Gassiot Committee Report.* (1947), p. 91, published by the Physical Society, London (1948).
38*b*. Meinel, A. B., *Pub. Astr. Soc. Pacif.*, **60**, 373 (1948).
38*c*. Meinel, A. B., *Trans. Amer. Geophys. Union*, **31**, 21 (1950).
38*d*. Forsyth, P. A., Petrie, W. and Currie, B. W., *Nature*, **164**, 453 (1949).
39. Vegard, L., *Erg. der Exakt. Naturwiss.*, **17**, 229 (1938).

40. Vegard, L., *Geofys. Pub.*, **9**, No. 11 (1932); *Geofys. Pub.*, **10**, No. 4 (1933); Vegard, L. and Harang, L., *Geofys. Pub.*, **10**, No. 5 (1933); Vegard, L. and Tönsberg, E., *Geofys. Pub.*, **11**, No. 16 (1937).
41. Störmer, C., *Nature*, **139**, 584 (1937).
42. Störmer, C., *Zeit. Geophys.*, **5**, 177 (1929); *Nature*, **124**, 263 (1929).
43. Vegard, L., *Geofys. Pub.*, **12**, No. 14 (1944); **16**, No. 1 (1945); see also *Gassiot Committee Report* (1947), p. 82, published by the Physical Society, London (1948).
44. Vegard, L., *Geofys. Pub.*, **12**, No. 14 (1944).
45. Barbier, D., *Contributions de l' Institut d' Astrophysique*, Series B, No. 31 (1949).
46. Tcheng Mao-Lin and Dufay, J., *Cahiers de Physique*, No. 8, p. 51 (1942).
47. Vegard, L., *Zeit. Geophys.*, **6**, 42 (1930); **7**, 196 (1931); Vegard, L. and Tönsberg, E., *Nature*, **137**, 778 (1936).
48. Darrow, K. K., *Electrical Phenomena in Gases*, Baillière, Tindall and Cox (London), p. 48 (1932).
49. Arnot, F. L., *Collision Processes in Gases* (Methuen), p. 89 (1913).
50. Rüchardt, E., *Ann. Phys., Lpz.*, **12**, 601 (1932).
51. Richardson, R. S., *Trans. Amer. Geophys. Union*, p. 558 (1944); Brück, H. A. and Ruttlant, F., *Mon. Not. Roy. Astr. Soc.*, **106**, 130 (1946).
51a. Meinel, A. B., *Astrophys. Jour.*, **113**, 50 (1951); Gartlein, G. W., *Nature*, **167**, 277 (1951).
52. Rutherford, E., *Radioactivity and Radioactive Substances*, Cambridge University Press, p. 87 (1913).
52a. *Ibid.*, p. 143.
53. Vegard, L., *Terrestrial Magnetism and Electricity*, Vol. VIII of series 'Physics of the Earth', McGraw-Hill, p. 609 (1939).
54. Vegard, L., *Ann. d. Phys.*, **50**, 853 (1916).
55. Vegard, L., *Phil. Mag.*, **42**, 47 (1921).
55a. Vegard, L., *Terrestrial Magnetism and Electricity*, Vol. VIII of series 'Physics of the Earth,' McGraw-Hill, p. 607 (1939).
55b. Vegard, L., *Geofys. Pub.*, **10**, No. 4 (1933).
56. Mitra, S. K., *Active Nitrogen—A New Theory* (1945), Indian Association for the Cultivation of Science, Jadavpur, Calcutta. An account of the older theories is to be found in pp. 41–44.
57. Bates, D. R., Massey, H. S. W. and Pearse, R. W. B., 'Emission Spectra of the Night Sky and Aurorae', *Gassiot Committee Report* (1947) p. 97, published by the Physical Society, London (1948).
58. Mitra, S. K., *Science and Culture (Calcutta)*, **9**, 46 (1943-44); *Nature*, **155**, 786 (1945). Ghosh, S. N., *Proc. Nat. Inst. Sci. (India)*, **9**, 301 (1943); **12**, 405 (1946).
59. Lozier, W. W., *Phys. Rev.*, **46**, 268 (1934).
60. Vier, D. T. and Mayer, F. E., *Jour. Chem. Phys.*, **12**, 28 (1944).
61. Bates, D. R., Buckingham, R. A., Massey, H. W. W. and Unwin, J. J., *Proc. Roy. Soc. A*, **170**, 322 (1939).
62. Bates, D. R. and Massey, H. S. W., *Phil. Trans. Roy. Soc. A*, **239**, 269 (1943).
63. Chapman, S., *Phil. Mag.*, **23**, 657 (1937).
64. Mitra, S. K., *Nature*, **167**, 897 (1951).
65. Gaydon, A. G., *Nature*, **153**, 407 (1944).
65a. Herzberg, G. and Herzberg, L., *Nature*, **161**, 283 (1948).
66. Nicolet, M., 'Emission Spectra of the Night Sky and Aurorae', *Gassiot Committee Report* (1947), p. 36, published by the Physical Society, London (1948).
67. Störmer, C., *Nature*, **142**, 1034 (1938).
68. Vegard, L., *Nature*, **124**, 947 (1929).
69. Slipher, V. M., *Mon. Not. Roy. Astr. Soc.*, **93**, 657 (1933).
70. Elvey, C. T., *Rev. Mod. Phys.*, **14**, 140 (1942).